

PRONÓSTICO DE VIDA ÚTIL RESTANTE EN RODAMIENTOS, CON BASE EN DATOS  
DE VIBRACIONES Y SISTEMAS DE INFERENCIA ESTOCÁSTICA CON DEGRADACIÓN  
NO LINEAL.

DIEGO FERNANDO MUÑOZ TORRES

UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
FACULTAD DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA, FÍSICA Y CIENCIAS DE  
LA COMPUTACIÓN  
MAESTRÍA EN INGENIERÍA ELÉCTRICA  
PEREIRA  
2016

PRONÓSTICO DE VIDA ÚTIL RESTANTE EN RODAMIENTOS, CON BASE EN DATOS  
DE VIBRACIONES Y SISTEMAS DE INFERENCIA ESTOCÁSTICA CON DEGRADACIÓN  
NO LINEAL.

Diego Fernando Muñoz Torres

Trabajo de grado para optar al título de

Magister en Ingeniería Eléctrica

Director

M.Sc Mauricio Holguín Londoño

UNIVERSIDAD TECNOLÓGICA DE PEREIRA  
FACULTAD DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA, FÍSICA Y CIENCIAS DE  
LA COMPUTACIÓN  
MAESTRÍA EN INGENIERÍA ELÉCTRICA  
PEREIRA  
2016

Nota de aceptación

---

---

---

---

---

---

---

Presidente del jurado

---

Jurado

---

Jurado

Pereira, Febrero 2016

## **DEDICATORIAS**

Dedico este trabajo a mis padres, que en todo momento me apoyan sin esperar nada a cambio, a mis hermanos, que han sido mis mejores amigos y consejeros durante la vida y con los cuales cuento y contaré siempre en momentos difíciles.

## **AGRADECIMIENTOS**

Agradezco a mis padres y hermanos por todo el apoyo que me han brindado durante mi proceso de formación. A mi papá le agradezco los buenos consejos que me ha dado con respecto a tomar decisiones al futuro. A mi mamá le debo que me enseñara los valores morales para ser una persona humilde, honesta y de buen corazón al momento de trabajar, estudiar o hacer cualquier actividad en la vida, pues solo los triunfos alcanzados por el propio esfuerzo perduran por siempre. A mis hermanos les agradezco que aguantaran mi mal genio, buenos y malos ratos, apoyaran mis ideas al futuro. Siempre podemos cambiar para mejorar.

Agradezco a mis amigos, que siempre me motivan e impulsan para que saquemos nuestras metas laborales y académicas adelante a pesar de las dificultades financieras o personales que se nos presenten. Siempre puedo contar con ellos en los buenos o malos momentos.

Agradezco al ingeniero Mauricio Holguín, por la confianza, paciencia, amistad e instrucción a lo largo de este proceso formativo y por haber seguido trabajando conmigo a pesar de los muchos contratiempos que he tenido en esta difícil época de mi vida. Le agradezco enormemente por haber sido un ejemplo a seguir tanto por sus conocimientos y convicción al momento de aplicar sus conocimientos, como al mismo tiempo enseñarme a ser una persona de muchas virtudes morales y humana.

Agradezco a todas las buenas personas que he conocido durante este proceso formativo, porque aunque no tuve la suerte de conocernos bien, he aprendido muchas cosas buenas de todos.

## TABLA DE CONTENIDO

TABLA DE CONTENIDO .....	1
LISTA DE FIGURAS.....	4
LISTA DE TABLAS.....	5
PARTE I	
PRELIMINARES .....	6
DEFINICIÓN DEL PROBLEMA.....	7
JUSTIFICACIÓN .....	8
OBJETIVOS.....	9
Objetivo general.....	9
Objetivos específicos.....	9
PARTE II	
MATERIALES Y MÉTODOS .....	10
CAPÍTULO 1	
MARCO REFERENCIAL .....	11
1.1. ANTECEDENTES.....	11
1.2. MARCO CONCEPTUAL.....	12
CAPITULO 2	
CREACIÓN DE LOS LIBROS DE CÓDIGOS. ....	13
2.1. CONSTRUCCIÓN DE LOS LIBROS DE CÓDIGOS. ....	13
2.2. ANÁLISIS DE SEÑALES.....	14
2.2.1. Extracción de los Coeficientes Cepstrales en la frecuencia de Mel (MFCC) .....	15
2.3. FUZZY K-MEANS.....	17
2.4. SPECTRAL CLUSTERING .....	19
2.4.1. Descripción de la composición Espectral de una Matriz de similaridad.....	19
2.4.2. Algoritmo de una Matriz de Similaridad .....	20
CAPITULO 3	
MÉTODOS DE CLASIFICACIÓN .....	21
3.1. PROCESOS GAUSSIANOS.....	21

3.1.1.	Acondicionamiento de las señales de Vibración .....	22
3.1.2.	Configuración del proceso Gaussiano.....	22
3.1.3.	Implementación del Proceso Gaussiano.....	23
3.2.	MODELOS OCULTOS DE MARKOV HMM.....	24
3.2.1.	Observaciones de Densidad Continua en HMM.....	26
3.2.2.	Mezclas Gaussianas.....	27
3.3.	TÉCNICAS ADICIONALES PARA LA IMPLEMENTACIÓN DE LOS PROCESOS GAUSSIANOS Y LOS HMM.....	28
3.3.1.	Normalización Estadística .....	28
3.3.2.	Técnicas de validación.....	28
3.3.3.	Escalamiento .....	30
3.3.4.	Múltiples secuencias de observación.....	31
3.3.5.	Efectos de datos insuficientes en el entrenamiento .....	32
CAPITULO 4		
	ADQUISICIÓN DE LAS SEÑALES DE VIBRACIÓN .....	33
	Asignación de severidad por falla en rodamientos.....	34
PARTE III		
	RESULTADOS Y CONCLUSIONES.....	36
CAPITULO 5		
	RESULTADOS .....	37
5.1	DISEÑO DE LA METODOLOGÍA PARA EL ENTRENAMIENTO DEL SISTEMA DE DETECCIÓN DE FALLAS.....	37
5.1.1	Metodología para la extracción de características relevantes.....	37
5.1.2	Creación de los libros de códigos.....	38
5.1.3	Entrenamiento de los Procesos Gaussianos y los HMM para el sistema de detección y clasificación .....	39
5.2	DISCUSIÓN ACERCA DEL SISTEMA DE DETECCIÓN Y CLASIFICACIÓN DE FALLAS.....	40
5.2.1	Resultados obtenidos para los MHMM. ....	41
5.2.2	Resultados obtenidos para los GP.....	43
5.2.3	Discusión de resultados.....	46
5.3	PLANTEAMIENTO DEL SISTEMA DE DIAGNOSTICO Y PRONOSTICO DE FALLAS CON BASE EN LOS RESULTADOS OBTENIDOS POR LOS CLASIFICADORES.....	48
CAPITULO 6		
	CONCLUSIONES Y TRABAJOS FUTUROS .....	50

APÉNDICE A	
DESCRIPCIÓN MATEMÁTICA DEL ALGORITMO DE ENTRENAMIENTO .....	56
APÉNDICE B	
SOLUCIÓN A LOS TRES PROBLEMAS BÁSICOS EN LOS HMM .....	58



## LISTA DE FIGURAS

Figura 1. Diagrama de Flujo para el algoritmo de K-Means Clustering.....	14
Figura 2. Diagrama de bloques para el cálculo de los MFCC.....	15
Figura 3. Banco de Filtros en la escala de Mel.....	16
Figura 4. Mezcla de 3 distribuciones gaussianas.....	27
Figura 5. Sistema de pruebas a rodamientos.....	34
Figura 6. Tiempo de entrenamiento y validación. ....	47
Figura 7. Resultados de la práctica. ....	47

## LISTA DE TABLAS

Tabla 1. Resultados con los datos del Fuzzy K-Means sobre el clasificador MHMM .....	41
Tabla 2. Resultados con los datos originales sobre el clasificador MHMM.....	42
Tabla 3. Resultados con los datos del Spectral Clustering sobre el clasificador MHMM.....	43
Tabla 4. Resultados con los datos del Fuzzy K-Means sobre el clasificador GP .....	44
Tabla 5. Resultados con los datos originales sobre el clasificador GP .....	45
Tabla 6. Resultados con los datos del Spectral Clustering sobre el clasificador GP	46

**PARTE I**  
**PRELIMINARES**

## DEFINICIÓN DEL PROBLEMA

Entre los principales problemas que se dan al momento de calcular la vida útil restante (RUL, del inglés Remaining Useful Life) está la dificultad para tener una base de datos que contenga toda la información del ciclo de vida útil de la máquina y el alto costo computacional que presenta el entrenamiento de un sistema de clasificación para realizar inferencia del RUL. Para dar solución a estos inconvenientes se han aplicado múltiples técnicas como el entrenamiento de los parámetros que describen el RUL de un activo por medio de las redes neuronales, las cuales han fracasado por sobre-entrenarse [1-3, 4]; también se ha realizado inferencia del RUL por medio de la combinación de señales provenientes de múltiples sensores y conocimiento empírico integrado mediante filtros de partículas [5], los cuales han dado buenos resultados pero dependen mucho de un conocimiento a priori del sistema, así como de una base de datos extensa [3, 6].

Por tales razones, los investigadores han empezado a explorar nuevos campos, en los cuales se encuentra la aplicación de los Modelos Ocultos de Markov (HMM, del inglés Hidden Markov Models) para el análisis de fallas en rodamientos [7-15], la aplicación de los procesos Gaussianos para la estimación del RUL en bienes mediante el conocimiento de señales de degradación [15-26], que han traído consigo mejoras en el mantenimiento predictivo.

Sin embargo, todavía no se cuenta con una metodología que permita mediante el uso de HMM diseñar un clasificador robusto del estado de falla del rodamiento y tampoco se cuenta con una metodología que permita inferir el comportamiento de la falla a futuro para los rodamientos, al tener solamente acceso a bases de datos incompletas por solo cubrir parte del ciclo de vida de la máquina [27-33].

Se toma como tema de investigación la inferencia del RUL para rodamientos, mediante el uso de HMM en sistemas dinámicos considerando la degradación no lineal entre los diferentes estados de falla, debido a la importancia que representan los rodamientos en las máquinas rotativas, y también como un aporte en la generación de conocimiento en el área de la automatización del mantenimiento para el desarrollo regional.

## JUSTIFICACIÓN

En la actualidad, se presta particular importancia al hecho de conocer con exactitud el tiempo de vida útil restante de un sistema industrial, sin embargo para un enfoque práctico la importancia recae sobre el hecho de conocer el estado actual de desgaste, o falla, presente en los equipos o herramientas que conforman estos sistemas industriales [1], ya que el buen funcionamiento en conjunto de estos garantiza el buen desempeño del sistema, y permite mantener los estándares de producción requeridos. En la industria, las máquinas giratorias son ampliamente utilizadas, desde labores de mantenimiento hasta tareas complejas como la automatización de sistemas. La experiencia muestra que la mayoría de las fallas en estas máquinas se presenta en los cojinetes de los rodamientos, los cuales al presentar fallas generan paros [2]. Para inferir el RUL de los rodamientos se usan técnicas como el análisis por redes bayesianas dinámicas, las cuales proporcionan una buena estimación del RUL, sin embargo, éstas exigen un alto costo computacional, además de requerir bases de datos con la historia completa de la vida útil del rodamiento [3, 4, 6]. También se estima el RUL mediante modelos de la integración de múltiples bases de datos usando filtros estocásticos que entregan una estimación del RUL, pero igual exigen un alto costo computacional, además de que se integra un gran error al unir múltiples bases de datos tomadas por diferentes sensores [13, 14].

Por tales limitaciones, se plantea una metodología basada en la estimación del RUL haciendo uso de los Modelos Ocultos de Markov, los cuales son usados en aplicaciones con señales cuasi-estacionarias como es el caso de clasificadores de señales sonoras como el habla humano [7], y en donde se puede asumir que las señales de las vibraciones generadas por un rodamiento se pueden aproximar a dicho tipo de señales [8]. También se plantea el uso de los procesos Gaussianos como un método innovador que ataca rotundamente las metodologías que exigen un alto costo computacional como las mencionadas en [13, 14], ya que los procesos Gaussianos al contrario de estos, tiene un costo computacional mejor y se adaptan fácilmente a cualquier tipo de señal sin sobrentrenarse y permitiendo un buen seguimiento al comportamiento de cualquier señal. Lo cual permite inferir el estado posterior del proceso para así plantear una metodología de mantenimiento predictivo [9, 10]. Actualmente se ha explorado con estos métodos aplicándolos en otras áreas del conocimiento como el pronóstico del RUL en turbinas [11], en el modelamiento de las funciones de riesgo de bienes útiles [12] y en el monitoreo de la salud de componentes electrónicos [33-45].

Las metodologías anteriores conforman parte del estado actual del arte para el análisis de señales de vibraciones en rodamientos, además atienden al quehacer e interés de estudio al interior del Grupo de Investigación en Automática de la Universidad Tecnológica de Pereira, formando un punto de vista novedoso y con terreno por explorar.

## **OBJETIVOS**

### **Objetivo general**

Diseñar una metodología para la inferencia de la vida útil restante en rodamientos, mediante el uso de los Modelos Ocultos de Markov y Procesos Gaussianos.

### **Objetivos específicos**

- Diseñar un sistema para la extracción de características en la escala de MEL aplicable a señales de vibración provenientes de acelerómetros ubicados en rodamientos bajo falla. Se emplean bases de datos disponibles.
- Realizar tratamiento de las señales de datos para mejorar la calidad de las características de la señal, para lo cual se diseñan libros de códigos (discretización de la señal) mediante el uso de Fuzzy K-mean Clustering y Spectral Clustering.
- Diseñar un clasificador para las señales provenientes del rodamiento usando Mixturas de Modelos Ocultos de Markov.
- Diseñar un clasificador para las señales provenientes del rodamiento usando Procesos Gaussianos.
- Inferir el estado posterior del rodamiento mediante los resultados obtenidos mediante las máquinas entrenadas y los resultados obtenidos en la tarea de clasificación.
- Diseñar un sistema de diagnóstico para rodamientos en base al RUL.
- Plantear una metodología teórica del pronóstico en base a los resultados encontrados por los clasificadores Gaussianos y los Modelos Ocultos de Markov.

**PARTE II**  
**MATERIALES Y MÉTODOS**

# **CAPÍTULO 1**

## **MARCO REFERENCIAL**

### **1.1. ANTECEDENTES**

Actualmente, con el encarecimiento de la materia prima y el crecimiento de la competencia en el mercado mundial, la productividad en las compañías de manufacturación ha pasado de ser un aspecto relevante a ser el objetivo primario. La productividad es proporcional a la producción que se tenga, por lo cual, incrementar la disponibilidad puede lograr un incremento en la producción.

La razón indicada anteriormente, enfatizan la gran importancia de un monitoreo continuo, para conocer el estado de degradación de la máquina y evitar pérdidas por fallas presentes durante la producción y que no fueron asumidas por falta del desconocimiento del estado de la maquinaria.

Muchos investigadores se han enfocado en el desarrollo de sistemas confiables que permitan catalogar el estado de las máquinas, o en su defecto de las partes críticas de las mismas, como es el caso de los rodamientos en los motores eléctricos. El objetivo es prevenir fallas y permitir programar paros seguros de reparación y mantenimiento. Una manera de estudiar las fallas es a través del análisis de la Vida Útil Restante (RUL, del inglés Remaining Useful Life), permitiendo conocer la proyección del desgaste actual del rodamiento, debido a que el conocimiento del desgaste actual no es suficiente para tomar una decisión en las labores de mantenimiento. Se emplean los Modelos Ocultos de Markov (HMM, del inglés Hidden Markov Models) y los Procesos Gaussianos (PG), como metodologías para inferir el comportamiento del desgaste del rodamiento hasta su posible falla.

En este proyecto se diseñará un sistema para el pronóstico del RUL en rodamientos, con base en datos de vibraciones y los procesos Gaussianos, donde previamente se realiza caracterización de las señales de vibración con fines de discriminar fallos. Los HMM permiten crear grupos que se encuentran inmersos dentro de una base de datos conformada por observaciones de las cuales solo se conocen algunas características, el punto clave de usar estos modelos radica en que no es necesario conocer el historial de fallas total de la máquina, sino un comportamiento actual y una base de datos estratificada, ya que en HMM los estados anteriores no afectan el valor del estado siguiente, propiedad que es bien vista para el estudio de señales cuasi-estacionarias, como es el caso de las vibraciones. Los procesos Gaussianos por otra parte presentan un modelo de bajo costo computacional con una alta expectativa a aplicaciones en tiempo real, que permiten una diferenciación más precisa entre estados de desgaste no lineal



Vecinos, debido a la manera de construcción la cual consiste un clasificador continuo a partir del compendio de múltiples clasificadores discretos Gaussianos.

## 1.2. MARCO CONCEPTUAL

**Proceso estocástico:** Un proceso estocástico es un proceso aleatorio que evoluciona de acuerdo con un parámetro que por lo general es el tiempo [9].

**Proceso sin memoria o Markoviano:** Si en un proceso estocástico, la transición a un estado siguiente solo depende del estado presente en que se encuentre el sistema o proceso, es decir, no importa el recorrido que hizo el proceso o sistema para llegar al estado presente, se dice que este proceso no tiene memoria o es “Markoviano” [9].

**Modelo Oculto de Markov:** Un modelo oculto de Markov o HMM (por sus siglas del inglés, Hidden Markov Model) es un modelo estadístico en el que se asume que el sistema a modelar es un proceso de Markov de parámetros desconocidos. El objetivo es determinar los parámetros desconocidos (u ocultos, de ahí el nombre) de dicha cadena a partir de los parámetros observables. Los parámetros extraídos se pueden emplear para llevar a cabo sucesivos análisis, por ejemplo en aplicaciones de reconocimiento de patrones [7].

**Fuzzy Clustering:** Es una técnica iterativa para el desarrollo de conglomerados, esta técnica al igual que muchas otras es derivada de la técnica original llamada C-means clustering, esta técnica de conglomerados, consiste en la elaboración de k grupos para una base de datos, en la cual cada muestra de la base de datos se le asigna un grupo correspondiente al cual tenga mayor pertenencia, esta técnica a diferencia de C-means no mide la distancia entre datos y grupos mediante una distancia euclidiana, lo cual ofrece mejores resultados en un análisis multidimensional.

**Spectral Clustering:** Es una técnica iterativa para el desarrollo de conglomerados. Esta técnica al igual que muchas otras es derivada de la técnica original llamada C-means, esta técnica de conglomerados, consiste en la elaboración de k grupos para una base de datos, en la cual cada muestra de la base de datos se le asigna un grupo correspondiente a la cual tenga la mayor pertenencia, esta técnica a diferencia de otras técnicas, mide la distancia entre grupos y datos mediante una distancia no euclidiana, lo cual ofrece mejores resultados en un análisis multidimensional.

## **CAPITULO 2**

### **CREACIÓN DE LOS LIBROS DE CÓDIGOS.**

Los libros de códigos se definen como el compendio de los datos filtrados y discretizados bajo uno de los criterios de agrupación, para este proyecto se hizo uso de dos criterios de agrupación para llevar a cabo la discretización de los datos, la cual tiene como objetivo mejorar los tiempos de ejecución del algoritmo sacrificando un porcentaje de la diversidad de valores encontrados en la base de datos.

A continuación se explica el proceso aplicado a la base de datos para crear los libros de códigos (o la discretización de los datos), la cual fue implementada mediante dos técnicas de agrupamiento, las cuales fueron Fuzzy K-Means y los Spectral Clustering.

#### **2.1. CONSTRUCCIÓN DE LOS LIBROS DE CÓDIGOS.**

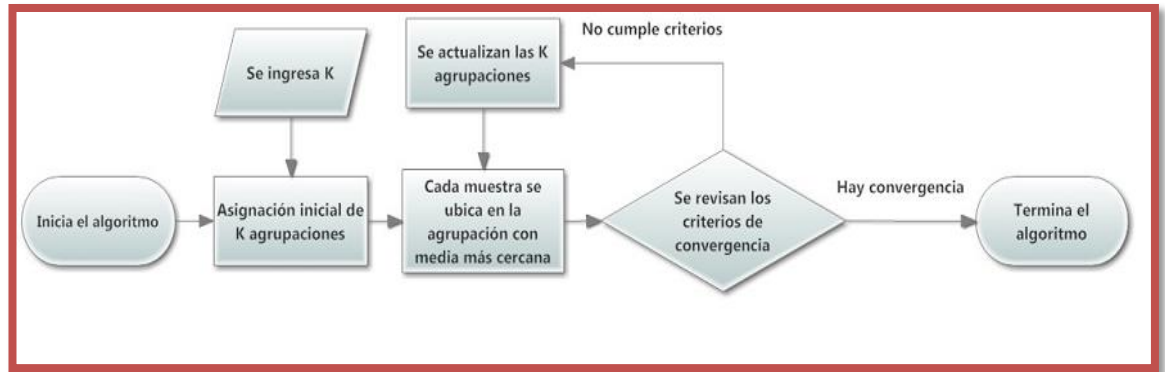
El algoritmo de entrenamiento empieza definiendo la cantidad de agrupaciones a utilizar, la cantidad de agrupaciones definidas en los algoritmos siempre será denominada por  $K$ , sin importar el método que se elija, donde cada una de estas agrupaciones representa un conjunto de datos.

El siguiente paso consiste en tomar cada dato del espacio muestral y asociarlo a la agrupación más cercana, este paso varía dependiendo del método de agrupamiento que se escoja, los cuales se mencionaron anteriormente. Finalmente se evalúa el desempeño de las asignaciones con respecto a una tolerancia asignada, si el sistema no converge se repiten los pasos anteriores [19-22].

1. Asignar la cantidad de agrupaciones  $K$ , utilizando un punto de partida inicial aleatorio.
2. Ubicar cada una de las muestras del espacio muestral en las  $K$  agrupaciones, en donde la muestra queda asignada a la agrupación con menor diferencia en su media.
3. Revisar los criterios de parada del proceso de asignaciones, estos criterios pueden ser número de iteraciones o por óptimo global, el óptimo global se asocia al hecho de que no se presenten cambios mayores entre la media de la agrupación actual con su anterior.

4. Si los criterios de parada no se cumplen, se actualizan las K agrupaciones y se regresa al paso 3.
5. Si los criterios de parada se cumplen, termina el proceso iterativo.

Figura 1. Diagrama de Flujo para el algoritmo de K-Means Clustering.



## 2.2. ANÁLISIS DE SEÑALES.

El análisis de señales consiste en el proceso de convertir una forma de onda cualquiera, en un conjunto de características conocidas con las cuales sea posible realizar un análisis matemático detallado. Para realizar este análisis es necesario captar de la señal dos conjuntos de parámetros, que corresponden a la frecuencia (tiempo) y la energía, con lo cual es posible modelar matemáticamente la señal [15].

Las señales pueden llegar a ser complejas, dependiendo directamente de la cantidad de información contenida. Un ejemplo de esto se ve en la voz humana, así para un hablante cualquiera se puede identificar su género, la edad, el mensaje que quiere transmitir, la emoción que transmite al momento de hablar, y si se cuenta con un registro de los hablantes, se podría identificar al mismo hablante [16].

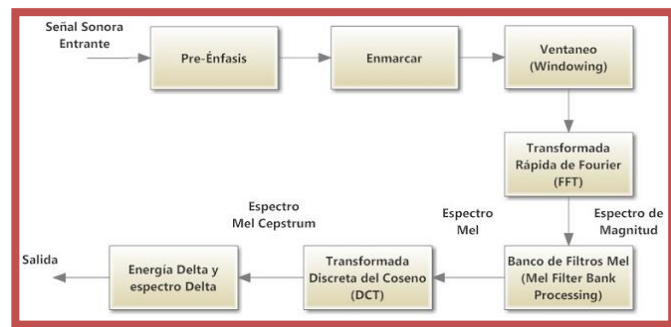
Varias técnicas han sido propuestas para reducir el desfase entre la prueba y los entornos de formación, en los cuales los métodos funcionan bien tanto para el espectro [17, 18] como para el dominio cepstral [19]. En primer lugar, la onda se convierte en una señal digital, con el fin de poder producir datos digitales que representen cada nivel de la señal en cada paso de tiempo, luego a estas señales digitalizadas se les extrae sus características usando coeficientes cepstrales en la frecuencia de Mel (MFCC), para crear así las funciones de voz correspondientes de la señal.

### 2.2.1. Extracción de los Coeficientes Cepstrales en la frecuencia de Mel (MFCC)

La extracción de características en señales de vibración es una tarea importante para mejorar el rendimiento en cualquier tarea de reconocimiento. La eficiencia en este proceso afecta la calidad de los modelos entrenados a la hora de llevarse a cabo una tarea de reconocimiento.

Los Coeficientes Cepstrales en la Frecuencia de Mel (MFCC) se basan en las percepciones auditivas humanas, las cuales no sobrepasan frecuencias de más de 1Khz. En otras palabras, los MFCC se basan en la variación conocida de ancho de banda crítica del oído humano con frecuencia entre Khz. Los MFCC tiene dos tipos de filtro que están espaciados de forma lineal en la frecuencia por debajo de 1000 Hz y con espaciamiento logarítmico sobre los 1000 Hz. Un tono subjetivo está presente en la escala de frecuencia de Mel, la cual permite capturar las características más importantes de la fonética en el habla. El proceso global de la MFCC se muestra en la figura 2 [20, 21]:

Figura 2. Diagrama de bloques para el cálculo de los MFCC.



Como se observa en el diagrama de bloques de la figura 2, para el procesamiento de la señal mediante los MFCC se requiere de 7 pasos computacionales con los cuales se hace el cambio de escala a una señal de voz:

#### Paso 1: Pre-Énfasis

Este paso procesa la señal a través de un filtro que hace hincapié en frecuencias más altas. Este proceso permite incrementar la energía de la señal de voz a altas frecuencias.

$$Y(n) = X(n) - 0.95X(n - 1) \quad (2.1)$$

Para el caso que se muestra, se considera una constante  $a = 0.95$ , lo que significa que un 95% de la muestra actual ha sido originada a partir de la muestra anterior.

#### Paso 2: Enmarcar

En este proceso, se organiza cada segmento de la señal de habla obtenida a partir de la conversión analógica a digital (ADC) en un pequeño marco de cierta longitud; se debe escoger adecuadamente estos marcos para cada señal, para el caso del habla se utiliza

un marco de referencia entre 20 a 40 mseg. Finalmente, lo que se obtiene es una señal contenida en muestras de N tramos, donde N representa la cantidad de marcos escogidos para la señal bajo estudio.

#### Paso 3: Ventaneo (Windowing)

En este paso se hace uso de las ventanas de Hamming, el proceso consisten en integrar todos los marcos de frecuencias cercanas creados anteriormente en un vector único que posee toda la información de la señal. Así, la salida que se tiene de la señal a través de la ventana, se muestra a continuación:

$$Y(n) = X(n) \times W(n) \quad (2.2)$$

$$W(n) = 0.54 - 0.46\cos[2\pi n/(N - 1)] \quad 0 \leq n \leq N - 1 \quad (2.3)$$

Donde:

N = Número de muestras en cada fotograma

Y(n) = Señal de salida

X(n) = Señal de entrada

W(n) = Ventana de Hamming,

#### Paso 4: Transformada rápida de Fourier (FFT)

Para convertir cada fotograma de N muestras del dominio del tiempo al dominio de la frecuencia, se hace uso de la transformada de Fourier, el procedimiento para este paso es el siguiente, hacer la convolución al pulso glotal U[n] y a la respuesta del tracto vocal de impulso h [n] en el dominio del tiempo. Esta declaración se apoya en la siguiente ecuación:

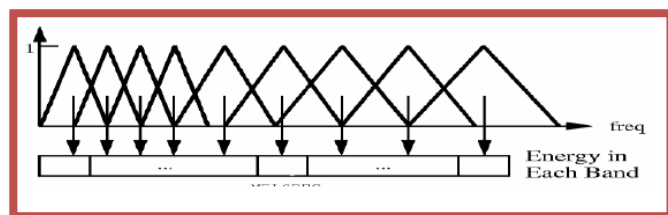
$$\begin{aligned} Y(w) &= \text{FFT} [H(t) * X(t)] \\ &= H(w) * X(w) \end{aligned} \quad (2.4)$$

En donde X(w), H(w) e Y(w) son la Transformada de Fourier de X (t), H (t) e Y (t), respectivamente.

#### Paso 5: Banco de filtros Mel para el procesamiento (Mel Filter Bank Processing)

El rango de frecuencias en el espectro de FFT es muy amplio y la señal no sigue una escala lineal, lo cual dificulta el almacenamiento de los datos de una manera correcta, por lo cual se hace necesario aplicar filtros. El banco de filtros de acuerdo con la escala de Mel se muestra en la figura 3.

Figura 3. Banco de Filtros en la escala de Mel.



En la figura 3 se muestra un conjunto de filtros triangulares, los cuales se utilizan para calcular una suma ponderada de componentes espectrales de modo que la salida del proceso se aproxime a una escala de Mel. Cada filtro de magnitud y frecuencia correspondiente posee una forma triangular e igual a la unidad de frecuencia central y disminuye linealmente hasta cero en una cantidad equivalente a la frecuencia central que hay entre los dos filtros adyacentes a él [22, 7]. Entonces, cada salida del filtro es la suma de sus componentes espectrales filtradas. La siguiente ecuación se utiliza para calcular la frecuencia de Mel en Hz:

$$F(\text{Mel}) = [2595 * \log_{10}[1 + f/700]] \quad (2.5)$$

#### Paso 6: Transformada Discreta del Coseno (DCT)

Este es el proceso que se lleva a cabo para convertir el espectro de registro en escala de frecuencia de Mel al dominio del tiempo, lo que se logra mediante la Transformada Discreta del Coseno (DCT). El resultado de la conversión de frecuencia se llama coeficiente Mel Cepstrum. Al conjunto de coeficientes Mel cepstrum se les denomina vectores acústicos. Por lo tanto, cada expresión de entrada se transforma en una secuencia del vector acústico.

#### Paso 7: Energía Delta y Espectro Delta

En cualquier otro tipo de señal cuasi-estacionaria, se pueden presentar cambios bruscos en sus marcos. Por lo tanto, existe la necesidad de añadir funciones relacionadas con el cambio en las características cepstrales a lo largo del tiempo, para esto se dispone de un conjunto de 13 funciones delta o también llamadas funciones de velocidad (12 funciones cepstrales más una de energía), y de 39 características de triángulo doble o también llamadas características de aceleración las cuales se deben agregar al análisis. Cada una de las 13 funciones delta representa el cambio entre los fotogramas. Cada una de las 39 funciones de triángulo doble representa el cambio entre marcos de las características de delta correspondientes, esta se describe matemáticamente como se muestra a continuación.

$$d(t) = \frac{C(t+1) - c(t-1)}{2} \quad (2.6)$$

### 2.3. FUZZY K-MEANS

El algoritmo de Fuzzy K-means clustering es implementado mediante varios pasos, entre estos pasos primero se debe escoger un punto inicial para los centros de los núcleos, este paso se puede hacer mediante una división de distancias equidistante del espacio muestral, con lo cual el algoritmo empezara a ordenar cada dato con respecto al núcleo más cercano, el segundo paso que realiza el algoritmo es actualizar la posición de los centroides, minimizando la función de costos, en este paso los centroides son ubicados

en la posición más cercana a los grupos de datos que al ser evaluados dieron el mejor valor de probabilidad de pertenencia a ese grupo[8, 16].

### Algoritmo de Fuzzy K-Means

Sea  $X = \{X_1, X_2, \dots, X_n\}$  un conjunto de objetos de tamaño  $n$ . Los objetos  $X = \{X_{i1}, X_{i2}, \dots, X_{in}\}$  están caracterizados por ser un conjunto compuesto por  $m$  características (variables o atributos). El algoritmo de Fuzzy K-means o el algoritmo de c-medias difusas [8, 16], busca particionar los datos  $X$  en  $K$  grupos, minimizando la función objetivo  $F$ , la cual posee variables desconocidas las cuales se nombran  $U$  y  $Z$ , la función  $F$  se expresa a continuación:

$$F(U, Z) = \sum_{l=1}^k \sum_{i=1}^n u_{il}^\alpha D(X_i, Z_l) \quad (2.7)$$

Sujeta a:

$$\begin{cases} \sum_{l=1}^k u_{l,i} = 1, & 1 \leq i \leq n \\ 0 \leq u_{l,i} \leq 1, & 1 \leq i \leq n, \quad 1 \leq l \leq k \\ 0 \leq \sum_{i=1}^n u_{l,i} < n, & 1 \leq l \leq k \end{cases} \quad (2.8)$$

Donde:

$U$  es una matriz de miembros de tamaño  $n \times k$  y  $\alpha \in [1, \infty)$  son el peso de los parámetros.

$Z = \{Z_1, Z_2, \dots, Z_k\}$  es el conjunto de  $k$  medias que conforman los centros de los  $k$  núcleos.

$D(X_i, Z_l)$ , se denota como la distancia de dis-similaridad medida entre la característica  $i$  al núcleo  $l$  en la  $j$ -ésima variable, usualmente se define como se muestra a continuación:

$$D(X_i, Z_l) = \sum_{j=1}^m (x_{i,j} - z_{l,j})^2. \quad (2.9)$$

Al abordar el problema de optimización se encuentra que este se puede resolver mediante la minimización de los dos siguientes problemas:

Problema  $P_1$ : se fija  $Z = \hat{Z}$ , esto ayuda a reducir el problema que se presenta en  $P(U, \hat{Z})$ ;  
 Problema  $P_2$ : Se fija  $U = \hat{U}$ , esto ayuda a reducir el problema que se presenta en  $P(\hat{U}, Z)$ ;

El problema  $P_1$  se resuelve con

$$U_{il} = \begin{cases} 1, & \text{Si } D_{il} = 0 \\ 0, & \text{Si } D_{ih} = 0, h \neq l \\ 1 / \sum_{h=1}^k \left( \frac{D_{il}}{D_{ih}} \right)^{\frac{1}{\alpha-1}}, & \text{Si } D_{ih} \neq 0 \text{ para } 1 \leq h \leq k. \end{cases} \quad (2.10)$$

Para  $1 \leq l \leq k$  y  $1 \leq i \leq n$ .

Ahora para solucionar el problema  $P_2$  usamos

$$Z_{l,j} = \frac{\sum_{i=1}^n u_{i,l} x_{i,j}}{\sum_{i=1}^n u_{i,l}} \text{ para } 1 \leq l \leq k \text{ y } 1 \leq j \leq m \quad (2.11)$$

## 2.4. SPECTRAL CLUSTERING

Los métodos espectrales han sido introducidos para el análisis de datos conglomerados para su aplicación en diversos campos. En particular los conglomerados espectrales son técnicas que hacen referencia a la estructura propia de la matriz de similaridad, en donde la pertenencia de un punto o una variable está asociada al núcleo con el cual comparta más similaridad [8, 16].

### 2.4.1. Descripción de la composición Espectral de una Matriz de similaridad.

Los algoritmos de conglomerados espectrales están compuestos de tres pasos fundamentales. El primer paso es construir la matriz de similaridad de los datos. La computación de este paso generalmente requiere que los datos estén normalizados. El segundo paso consiste en computar la matriz de Laplacianos, en el cual se debe asignar el valor de núcleos deseados. Los núcleos formados por los datos proyectan el espacio propio de los datos, debido a esto la mayor cantidad de núcleos posibles es equivalente a la cantidad de valores propios de la matriz. El tercer paso consiste en archivar los



valores propios que fueron reconocidos por los  $k$  núcleos en una matriz de relación o de pertenencia.

Este método tiene un poder alto de aprendizaje ya que tiene la habilidad de reconocer núcleos no convexos y evitar caer en mínimos locales con facilidad. Como puede suceder al realizar las primeras iteraciones.

#### 2.4.2. Algoritmo de una Matriz de Similitud

Se tiene un conjunto de puntos  $C = \{C_1, C_2, \dots, C_n\}$  el cual se quiere clasificar en  $k$  clases. Construir la matriz de similitud  $A \in R^{n \times n}$ , con lo cual se define  $A_{ij}$ , en donde cada punto del conjunto de datos  $C_i$  tiene un valor entre 0 y 1 con respecto a cada núcleo.

1. Hallar la matriz diagonal  $D \in R^{n \times n}$  en donde  $D_{ii} = \sum_{m=1}^n A_{im}$ , y definir la matriz de similitud  $L = D^{-1/2}AD^{-1/2}$ .
2. Extraer los  $k$  valores propios principales  $\{X_1, \dots, X_k\}$  de  $L$ , y con estos formar la matrix  $X$  que contenga todos los valores propios en las columnas.
3. Calcular la matriz  $Y$  normalizando cada línea de  $X$ :  $Y_{ij} = X_{ij} / \sum_{m=1}^n X_{im}^2$ .
4. Considerando que cada punto de  $C_i \in C$  es transformado en  $C'_i \in R^k$  donde las cordenadas son  $\{Y_{im}, m \in \{1, \dots, k\}\}$ , se aplica la transformada de los  $k$ -means clustering para transformar cada punto al centroide más cercano.
5. Finalmente se asigna cada punto original  $C_i$  a su clase correspondiente según la transformada  $C'_i$ .

## CAPITULO 3

### MÉTODOS DE CLASIFICACIÓN

Los modelos continuos proporcionan un clasificador del estado actual de desgaste de la máquina, así como una función que describe en gran medida el comportamiento de una maquinaria ligado a la cantidad de información que se le presente mediante datos tomados. Los procesos Gaussianos y los Modelos Ocultos de Markov permiten crear grupos que se encuentran inmersos dentro de una base de datos conformada por observaciones. Estos permiten trabajar sobre datos con comportamiento no lineal, siempre y cuando la señal sea continua o al menos sea continua por tramos (señal cuasi-estacionaria). En este documento se presenta el entrenamiento de un sistema para clasificación continua, que evita los errores de clasificación presentes entre dos estados vecinos, haciendo uso de los procesos Gaussianos. El objetivo final es formar un clasificador continuo a partir del compendio de múltiples clasificadores discretos Gaussianos, para lo cual se hace un pre procesamiento de la señal y extracción de características por Coeficientes Cepstrales.

Ambos modelos tienen importantes aplicaciones en confiabilidad. La simplicidad y poder analítico de ambos métodos para entrenar modelos de detección y clasificación, los han hecho reconocidos en el estudio de todo tipo de ondas cuasi-estacionarias como el habla humano o el ruido generado por una máquina rotativa como se demostrará en capítulos posteriores en donde se planteará una metodología para este tipo de análisis.

#### 3.1. PROCESOS GAUSSIANOS.

Se define un proceso Gaussiano sobre la función  $a(x)$  y luego se transforma con una función logística sigmoideal  $\sigma(a)$ , así se obtiene un proceso estocástico. La distribución de probabilidad sobre la variable de destino  $t$  está entonces dada por la distribución de Bernoulli, ecuación 3.1.

$$p(t/a) = \sigma(a)^t(1 - \sigma(a))^{1-t} \quad (3.1)$$

Se denota las entradas de formación establecidas por  $X_1, \dots, X_n$  con las correspondientes variables objetivo observadas y guardadas en  $t = (t_1, \dots, t_n)^T$ . Se considera un solo punto de prueba  $x_{n+1}$  con valor objetivo  $t_{n+1}$ . El objetivo del proceso Gaussiano es determinar la distribución predictiva  $p(t_{n+1}/t)$ . Para solucionar el posterior se introduce un proceso Gaussiano.

El Prior del proceso Gaussiano para  $a_{n+1}$  toma la forma mostrada en la ecuación 3.2.

$$p(a_{n+1}) = N(a_{n+1}/0, C_{n+1}) \quad (3.2)$$

A diferencia del caso de regresión, la matriz de covarianza no incluye un término de ruido porque supone que todos los puntos de datos de entrenamiento están etiquetados correctamente. Sin embargo, por razones numéricas es conveniente introducir un término de ruido gobernado por un parámetro  $v$ , que asegura que la matriz de covarianza sea definida positiva. Por lo tanto la matriz de covarianza  $C_{n+1}$  tiene elementos dados por la ecuación 3.3.

$$C(X_n, X_m) = k(X_n, X_m) + v\delta_{nm} \quad (3.3)$$

donde  $k(x_n, x_m)$  es cualquier función del núcleo semidefinida positiva, el valor de  $v$  está normalmente fijado por adelantado.

Para los problemas en donde las clases se comparan, es suficiente predecir  $p(t_{n+1} = 1/T_n)$  porque el valor de  $p(t_{n+1} = 0/T_n)$  viene dado por  $1 - p(t_{n+1} = 1/t_n)$ . La distribución predictiva requerida viene dada por la ecuación 3.4.

$$p(t_{n+1} = 1/t_n) = \int p(t_{n+1} = 1/a_{n+1}) p(a_{n+1}/t_n) da_{n+1} \quad (3.4)$$

Esta integral es analíticamente intratable, por lo que se puede aproximar utilizando métodos de muestreo. Alternativamente, se puede considerar técnicas basadas en una aproximación analítica, para más detalle se puede mirar la referencia [8].

### 3.1.1. Acondicionamiento de las señales de Vibración

En este paso, se discuten los criterios tomados para el cambio del dominio temporal de la señal al dominio tiempo-frecuencia representado en los coeficientes Cepstrales en escala Mel como se muestra en el capítulo 2. El primer paso es fijar los parámetros que se desean manipular, que para este caso son: cantidad de coeficientes Cepstrales por ventana de la señal, cantidad de filtros Mel a aplicar y finalmente el traslape entre ventanas. El tamaño de las ventanas se fija en 200ms, la idea de este paso es implantar una metodología que permita establecer heurísticas para el entrenamiento de un modelo clasificador de fallas, sin omitir información relevante de las señales originales y además optimizar los recursos informáticos.

### 3.1.2. Configuración del proceso Gaussiano.

Para la configuración del proceso Gaussiano, se procede como se describe a continuación.

Se eligen los hiperparámetros, de tal forma que no se escoja un mínimo local [8, 9]. El método de minimización a emplear es aproximación por Laplace, con el cual se minimiza el error para

inferir los parámetros del modelo. Así mismo, se establece la media de los parámetros y la covarianza como los valores iniciales de los parámetros del modelo. Esto se hace debido a que la función es no uni-modal.

Con los valores iniciales, se establece el cálculo aproximado del posterior, ecuación 3.5:

$$q(f/D) = \mathcal{N}(f/\mu, V) \quad (3.5)$$

La inferencia exacta sólo es manejable para probabilidades Gaussianas.

El kernel para poder medir distancias y hacer las comparaciones es el que expresa en la ecuación 3.6:

$$\begin{aligned} k_{\psi}(x, x') &= V[f(x)f(x')] = E[(f(x) - m(x))(f(x') - m(x')))] \\ k(x, x') &= \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}(x - x')^T(x - x')\right) \\ \psi &= \{\ln \ell, \ln \sigma_f\} \end{aligned} \quad (3.6)$$

Así, finalmente se tiene la expresión de la probabilidad como en la ecuación (3.7):

$$\text{likGauss}(t) = \frac{\exp(-(t - y)^2/2sn^2)}{\sqrt{2\pi sn^2}} \quad (3.7)$$

Donde “y” es la media y “sn” es la desviación estándar.

Con la anterior función se entrenan los parámetros del proceso Gaussiano que describe la base de datos en estudio.

Una vez entrenados los parámetros del proceso Gaussiano, se verifican los parámetros entrenados, usando el test de Bernoulli para pruebas binarias, con el cual se tiene el grado de pertenencia de cada señal a un respectivo grupo.

### 3.1.3. Implementación del Proceso Gaussiano.

Para la implementación del algoritmo, se recurre al uso de la aproximación por La Place, se elige la aproximación por Laplace para evadir una alta complejidad en el algoritmo y así minimizar el esfuerzo computacional requerido. La complejidad del algoritmo se puede definir mediante la matriz simétrica positiva.

$$B = I + W^{\frac{1}{2}} K W^{\frac{1}{2}}$$

Donde el costo computacional es  $O(n^2)$ , donde w es una matriz diagonal, B es la matriz de valores propios, B es la matriz de covarianzas.

### Algoritmo del proceso Gaussiano.

Entrada:  $K$  (matriz de covarianza),  $y$  ( $\pm 1$  objetivos),  $p(y/f)$  (Función de verosimilitud)

$f := 0$  (Inicialización)  
*repetir Iteración de Newton*  
 $W := -\nabla \nabla \log p(y/f)$  se evalua  $W$   
 $L := \text{cholesky}(I + W^{\frac{1}{2}} K W^{\frac{1}{2}})$   
 $b := Wf + \nabla \log p(y/f)$   
 $a := b - W^{\frac{1}{2}} L^T (L \setminus (W^{\frac{1}{2}} K b))$   
 $f := Ka$   
*hasta converger*  
 $\log q(y/X, \theta) := -\frac{1}{2} a^T f + \log p(y/f) - \sum_i \log L_{ii}$   
*regresar:*  $f := f(\text{pos modo}), \log q(y/X, \theta)$

### 3.2. MODELOS OCULTOS DE MARKOV HMM

Un modelo oculto de Markov (Hidden Markov Model HMM), es un modelo estadístico utilizado para modelar procesos de Markov de parámetros no observables u ocultos. Las cadenas de Markov generalmente asumen una cantidad finita de estados en donde cada estado es asociado a un evento físicamente observable. Cuando estos estados no se pueden definir de una manera directa, por la complejidad de las observaciones, es recomendable construir un modelo oculto de Markov, el cual, corresponde a un proceso estocástico doblemente inmerso, donde se posee una cadena de estados ocultos y en donde cada estado a su vez posee una matriz de emisión de símbolos con características propias [6,26-29].

Así los modelos ocultos de Markov se pueden caracterizar formalmente mediante los siguientes parámetros:

1.  $N$ , es el número de estados en el modelo. Los cuales normalmente están ocultos para la mayoría de las aplicaciones, estos estados generalmente están conectados entre sí de manera que se pueda ir de un estado oculto a otro (cadena ergódica), estos estados se denotan como  $S = \{S_1, S_1, \dots, S_N\}$ , y el estado en el tiempo  $t$  se denota como  $q_t$ .
2.  $M$ , es el número de símbolos diferentes de observaciones por estado, los cuales deben estar contenidos en un alfabeto discreto. Los símbolos de observación corresponden a

las salidas físicas del sistema que se quiere modelar. Se denotan los símbolos individuales como  $V = \{v_1, v_2, \dots, v_M\}$ .

3.  $A$ , es la matriz de probabilidad de transición de estado,  $A = \{a_{ij}\}$  donde

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N \quad (3.11)$$

En el caso de que para cualquier estado, se pueda llegar a otro estado en un paso, se tendrá que  $a_{i,j} > 0$  para todo  $i, j$ . de lo contrario se puede tener  $a_{i,j} = 0$  para uno o más pares.

4.  $B$ , es la matriz de probabilidad de emisión de los símbolos en el estado  $j$ .

$$B = \{b_j(k)\}, \text{ Donde}$$

$$b_j(k) = P[V_k \text{ en } t | q_t = S_j], \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad (3.12)$$

5.  $\pi$ , es el vector de estado inicial  $\pi = \{\pi_i\}$ , donde

$$\pi_i = P[q_1 = S_i], \quad 1 \leq i \leq N \quad (3.13)$$

Por otra parte si se tiene los valores de  $S$ ,  $V$ ,  $A$ ,  $B$  y  $\pi$  del modelo oculto de Markov se puede generar una secuencia de observaciones (donde cada observación  $O_t$ , es un símbolo de  $V$  y  $T$  es el número de observaciones en una secuencia).

$$O = O_1, O_2, \dots, O_T$$

El algoritmo para generar estas observaciones es el siguiente:

1. Se elige un estado inicial  $q_1 = S_i$ , el cual está contenido en el vector de estado inicial  $\pi$ .
2. Colocar  $t=1$ .
3. Escoger  $O_t = V_k$  de acuerdo a la matriz de emisión para el estado  $S_i$ .
4. Pasar al nuevo estado  $q_{t+1} = S_j$  de acuerdo a la matriz de probabilidades de transición para el estado  $S_i$ .
5. Colocar  $t = t + 1$ ; regresar al paso 3 Si  $t \leq T$ ; de lo contrario termina el proceso.

Finalmente por conveniencia, se denotan los modelos ocultos de Markov como se muestra en la ecuación 3.14.

$$\lambda = (A, B, \pi) \quad (3.14)$$

### 3.2.1. Observaciones de Densidad Continua en HMM

Toda la discusión que se ha llevado a cabo para los HMM, aplica solo para el caso en que se tienen observaciones de símbolos discretos contenidos en un alfabeto finito. Una manera de tratar las señales continuas ha sido cuantizándolas mediante un libro de códigos, lo cual permite la aplicación de los modelos descritos anteriormente a cambio de degradar la información en la señal. Existen otras formas de tratar las señales continuas sin degradar la información como pasa con la generación de los libros de código, esto se puede lograr asumiendo un arreglo de una o varias funciones de densidad de probabilidad (pdf) con el fin de modelar el comportamiento de los símbolos de las observaciones. Finalmente estas pdfs se pueden utilizar en los procesos de re-estimación, lo cual es propuesto como una mezcla finita. Comúnmente se hace uso de funciones de densidad simétrica log-cóncavas o elípticas, con la cual el modelo queda descrito de la siguiente forma [6, 33-35]:

$$b_j(\mathbf{O}) = \sum_{m=1}^M c_{jm} \mathfrak{N}[\mathbf{O}, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm}], \quad 1 \leq j \leq N \quad (3.15)$$

Donde  $\mathbf{O}$  es el vector que se usa para modelar,  $c_{jm}$  es el coeficiente de mezcla para la  $m$ -ésima mezcla en el estado  $j$  y  $\mathfrak{N}$  es cualquier densidad simétrica log-cóncava o elíptica,  $\boldsymbol{\mu}_{jm}$  es el vector de medias y  $\mathbf{U}_{jm}$  es la matriz de covarianzas para cualquier  $m$ -ésima componente de la mezcla en el estado  $j$ .

La ganancia de la mezcla debe satisfacer la siguiente restricción.

$$\sum_{m=1}^M c_{jm} = 1, \quad 1 \leq j \leq N \quad (3.16)$$

$$c_{jm} \geq 0 \quad 1 \leq j \leq N, \quad 1 \leq m \leq M \quad (3.17)$$

La fdp debe estar normalizada apropiadamente.

$$\int_{-\infty}^{\infty} b_i(x) dx = 1, \quad 1 \leq j \leq N \quad (3.18)$$

Finalmente se aplica los coeficientes de las mezclas de densidad a las fórmulas de re-estimación.

$$\bar{c}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k)}{\sum_{t=1}^T \sum_{k=1}^M \gamma_t(j, k)} \quad (3.19)$$

$$\bar{\boldsymbol{\mu}}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) \cdot \mathbf{O}_t}{\sum_{t=1}^T \gamma_t(j, k)} \quad (3.20)$$

$$\bar{\mathbf{U}}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) \cdot (\mathbf{O}_t - \bar{\boldsymbol{\mu}}_{jk})(\mathbf{O}_t - \bar{\boldsymbol{\mu}}_{jk})'}{\sum_{t=1}^T \gamma_t(j, k)} \quad (3.21)$$

$$\gamma_t(j, k) = \left[ \frac{\alpha_t(j) \beta_t(j)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)} \right] \left[ \frac{c_{jk} \mathfrak{N}(\mathbf{O}, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm})}{\sum_{m=1}^M c_{jk} \mathfrak{N}(\mathbf{O}, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm})} \right] \quad (3.22)$$

El término  $\gamma_t(j, k)$  presentado en la ecuación 3.22 es el caso general de la misma ecuación presentada en 3.15, para el cual la cantidad de mezclas corresponde a 1. Los casos compuestos por mezclas funcionan de la misma forma que los casos discretos.

### 3.2.2. Mezclas Gaussianas.

En la medición de datos de muchos procesos aleatorios, a menudo se presentan tendencias en su comportamiento a una distribución normal o gaussiana, lo que abre la posibilidad de que estos datos provengan de una mezcla de distribuciones gaussianas. Se describe entonces una distribución gaussiana como se muestra en la ecuación 3.23.

$$n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.23)$$

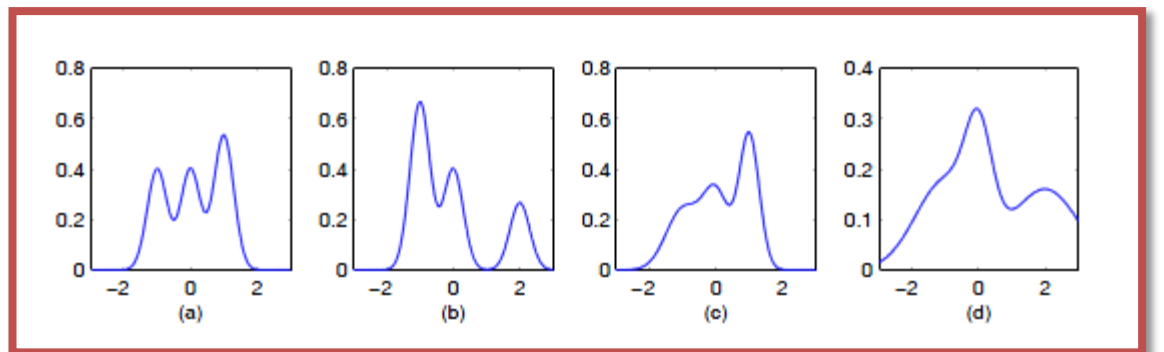
Donde  $\mu, \sigma$ , son respectivamente la media y la varianza de la fdp gaussiana. Ahora para el caso de tener una fdp compuesta de la mezcla de dos gaussianas, la representación matemática de esta mezcla viene dada por.

$$y(x, \mu, \sigma) = w_1 n_1(x, \mu, \sigma) + w_2 n_2(x, \mu, \sigma) \quad (3.24)$$

En la ecuación 3.24 el parámetro  $w$  representa el peso que se asocia a cada gaussiana respectivamente.

En el caso del análisis del ruido generado por las máquinas rotativas, el comportamiento de los datos se puede ajustar a una mezcla de distribuciones gaussianas, en la cual se debe elegir el número de mezclas adecuado para así ajustar lo mejor posible el comportamiento de la señal [36-41].

Figura 4. Mezcla de 3 distribuciones gaussianas.





En la figura 4 se muestra la mezcla de 3 distribuciones gaussianas, las cuales se presentan de cuatro maneras diferentes Figura 5(a), (b), (c), (d). Para este caso el conjunto de mezclas gaussianas puede ser representado de igual manera aclarando que se desconocen sus parámetros  $w$ ,  $\mu$  y  $\sigma$ . Se define  $\theta = \{\mu_j, \sigma_j^2\}$ , como la representación de una fdp gaussiana. La expresión general para cualquier cantidad de mezclas dada una señal se muestra a continuación.

$$P(y) = \sum_{j=1}^k w_j \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y-\mu_j)^2}{2\sigma_j^2}} \quad (3.25)$$

### 3.3. TÉCNICAS ADICIONALES PARA LA IMPLEMENTACIÓN DE LOS PROCESOS GAUSSIANOS Y LOS HMM

#### 3.3.1. Normalización Estadística

Se asocia con la acción de transformar una distribución cualquiera a una distribución normal. Corresponde a ajustar los datos de la distribución “inicial” a una distribución normal. En este caso se cambia la forma de la distribución original manteniendo la proporción de casos entre valores contiguos [42].

##### Normalización Zscore

En estadística la normalización Zscore sirve para ajustar los datos de las observaciones lo mejor posible a la media, dada la desviación estándar de la población [42].

La normalización Zscore se define:

$$y' = \frac{y-\mu}{\sigma} \quad (3.26)$$

Donde

$y$  Representa el dato de la población bajo ajuste.

$\mu$  Es la media de la población.

$\sigma$  Es la desviación estándar de la población.

#### 3.3.2. Técnicas de validación

Existen varias técnicas de validación, que se pueden utilizar a la hora de realizar un proceso de prueba a un sistema de software en un proyecto bajo estudio estadístico, entre ellas se pueden

encontrar la validación simple, la cual consiste en dividir el conjunto de muestras en dos grupos de manera aleatoria, también se tiene el método de validación cruzada o (rotation estimation en inglés) el cual es el comúnmente usado, ya que garantiza tanto la independencia entre los datos de entrenamiento con los de prueba al tiempo que asegura una adecuada partición de la base de datos.

### **Validación cruzada de K-iteraciones**

Este tipo de validación cruzada es la más simple, consiste en hacer K iteraciones en donde los datos de muestra se dividen en k subconjuntos. El sistema es entrenado con k-1 subconjuntos y el restante es utilizado para realizar las pruebas. Este proceso se realiza k veces, que es finalmente la cantidad de posibles combinaciones que se tienen de los datos. Este método es muy preciso ya que permite evaluar a partir de k combinaciones los datos de entrenamiento y de prueba, la desventaja de este método es que es muy lento en comparación al método de partición, el cual en esencia es el mismo método pero realizando una sola combinación [44, 45].

### **Validación cruzada simple**

Este método es una variación del método de validación cruzada de K iteraciones. En el cual, las muestras son divididas en dos subconjuntos únicamente, los cuales se denotan como d0 y d1, la elección de datos que conforman los grupos se hace de manera aleatoria, la única condición que se tiene es que ambos subgrupos queden conformados por igual cantidad de datos. La manera de entrenamiento que se realiza con este método es entrenar los modelos usando el subconjunto de datos d0 y probarlo con el subconjunto de datos d1, luego se realiza el mismo procedimiento pero entrenando con d1 y probando con d0 [44, 45]

Este método es mucho más rápido que el método de las k iteraciones pero tiene una desventaja, la cual consiste en que los datos pueden quedar o no contenidos en un grupo, o por el contrario puede presentarse que se repitan muchos datos en ambos grupos.

### **Validación cruzada por repetición aleatoria de sub muestras**

Este método consiste en dividir de manera aleatoria la base de datos en dos grupos, el de entrenamiento y el de validación. Los datos de entrenamiento son ajustados mediante los modelos entrenados, y la probabilidad del modelo es evaluada mediante los datos de entrenamiento. La ventaja de este método es que los grupos de entrenamiento y prueba no dependen del número de iteraciones. La desventaja de este método es que algunas observaciones pueden quedar excluidas de los conjuntos de entrenamiento y validación. Este método es muy aplicado en la simulación de monte carló [44, 45].

Finalmente la validación cruzada tiene como fin, estimar que tan afine es un modelo entrenado mediante un subconjunto de datos con respecto a otro subconjunto de datos, los cuales serán usados para evaluar el modelo o llevarse a cabo un test, estos datos de evaluación deben ser independientes de los datos de entrenamiento.

### Técnica de validación diseñada.

La técnica de validación diseñada para el proyecto de maestría es una combinación entre la validación cruzada simple y la validación por repetición aleatoria de sub-muestras, en la cual la base de datos se divide de manera aleatoria como se hace en la repetición aleatoria de sub-muestras y se hace K- veces como en la validación cruzadas de K-iteraciones, aunque para nuestro caso no se recurre a implementar todas las posibles k iteraciones sino una cantidad reducida, la cual se menciona en cada caso en la configuración de los modelos.

### 3.3.3. Escalamiento

El escalamiento es importante en las aplicaciones hechas con HMM cuando se posee una gran cantidad de símbolos de observaciones, que es lo que habitualmente se presenta dada la naturaleza de captación de las señales.

El algoritmo de Baum-Welch presenta problemas de convergencia cuando se tiene T demasiado grande en donde T representa el tamaño del conjunto de los símbolos de observaciones para una muestra. Ya que los valores que toma el algoritmo tienden a 0. La idea del escalamiento es evitar este contratiempo, el escalamiento consiste en escoger un conjunto de coeficientes escalares con los cuales se mantienen las probabilidades en una escala trabajable para el algoritmo. El coeficiente de escalamiento viene dado por [6, 25].

$$c_t = \frac{1}{\sum_{i=1}^N \alpha_t(i)} \quad (3.27)$$

Este coeficiente de escalamiento emplea sobre las variables forward y Backward del algoritmo de Baum-Welch, cumpliendo siempre que  $\sum_{i=1}^N c_t \alpha_t(i) = 1$ , al igual que  $\sum_{i=1}^N c_t \beta_t(i) = 1$  para todo  $1 \leq t \leq T$ . Las variables  $\alpha_t(i)$  y  $\beta_t(i)$  se calculan durante todo el proceso, por lo tanto el factor de escala total viene dado por:

$$c_\alpha(t) = \prod_{k=1}^t c_k \quad (3.28)$$

El mismo procedimiento se aplica para la variable Backward.

$$c_\beta(t) = \prod_{k=t}^K c_k \quad (3.29)$$

Finalmente lo que se obtiene al aplicar los factores de escalonamiento totales al algoritmo de Baum-Welch son las variables escaladas  $\alpha'_t(i)$ ,  $\beta'_t(i)$ ,  $\xi'_t(i, j)$ , las cuales quedan descritas con más claridad cómo se muestra a continuación.

$$\sum_{i=1}^N \alpha'_T(i) = c_\alpha(T) \sum_{i=1}^N \alpha_T(i) = c_\alpha(T) P(O|\lambda) \quad (3.30)$$

$$\xi'_t(i, j) = \frac{c_\alpha(t-1) \alpha_{t-1}(i) a_{ij} b_j(o_t) \beta_t(j) c_\beta(t)}{c_\alpha(T) \sum_{k=1}^N \alpha_T(k)} \quad (3.31)$$

Donde  $\xi'_t(i, j)$ , representa La probabilidad intermedia escalada.

De esta manera todos los coeficientes escalados se pueden utilizar sin inconvenientes en el algoritmo de EM, debido a que al finalizar el proceso de re-estimación serán canceladas, lo cual permite obtener los valores reales. Sin embargo para este proceso existe una excepción que viene dada al aplicarse los factores de escalamiento en la  $P(O|\lambda)$  para tal caso se recurre entonces a la siguiente expresión.

$$P(O|\lambda) = \frac{\sum_i \alpha'_T(i)}{c_\alpha(T)} \quad (3.32)$$

### 3.3.4. Múltiples secuencias de observación.

La necesidad de usar múltiples observaciones, nace de la dificultad latente de obtener un modelo confiable al ser entrenado con una sola observación, esto conduce generalmente a un modelo insuficientemente entrenado, que resulta inútil a la hora de evaluar la probabilidad a posteriori de las observaciones. Por lo tanto con el fin de poder entrenar un modelo lo suficientemente robusto para una aplicación, se recurre al uso de múltiples observaciones del mismo tipo pero independientes entre sí. La notación para las múltiples observaciones queda definida de la siguiente manera [1]:

$$\mathbf{O}^{(k)} = \{\mathbf{O}_1^k, \mathbf{O}_2^k, \dots, \mathbf{O}_{T_k}^k\} \quad (3.33)$$

Los parámetros de maximización quedan definidos de la siguiente manera:

$$P(O|\lambda) = \prod_{k=1}^k P(\mathbf{O}^{(k)}|\lambda) \quad (3.34)$$

$$= \prod_{k=1}^k P_k. \quad (3.35)$$

Actualizando el conjunto de ecuaciones para re-estimación obtenemos.

$$\bar{a}_{ij} = \frac{\sum_{k=1}^k \frac{1}{P_k} \sum_{t=1}^{T_{k-1}} \alpha_t^k(i) a_{ij} b_j(O_{t+1}^{(k)}) \beta_{t+1}^k(j)}{\sum_{k=1}^k \frac{1}{P_k} \sum_{t=1}^{T_{k-1}} \alpha_t^k(i) \beta_t^k(i)} \quad (3.36)$$

Y

$$\bar{b}_j(l) = \frac{\sum_{k=1}^k \frac{1}{P_k} \sum_{t=1}^{T_{k-1}} \alpha_t^k(i) \beta_t^k(i)}{\sum_{k=1}^k \frac{1}{P_k} \sum_{t=1}^{T_{k-1}} \alpha_t^k(i) \beta_t^k(i)} \quad (3.37)$$

No es necesario re-estimar la variable  $\pi_i$  para este caso.

### 3.3.5. Efectos de datos insuficientes en el entrenamiento

Otro problema que se presenta al realizarse aplicaciones con los GP y los HMM a sistemas reales es el efecto de datos insuficientes en el entrenamiento, este problema consiste en que al llevar a cabo la tarea de entrenamiento, la cantidad de estados propuesta es muy alta, o las observaciones tomadas son demasiado escasas, lo que provoca que el modelo entrenado no tenga un buen rendimiento a la hora de calcular  $P(O|\lambda)$ . Para este problema se trazan varias soluciones prácticas que en la mayoría de los casos pueden mejorar el rendimiento del modelo entrenado, estas son propuestas en [2, 6] y se mencionan a continuación.

La primera opción consiste en aumentar la cantidad de observaciones si es posible, esta opción se puede tomar si es permitido captar más datos del sistema. La segunda opción consiste en disminuir la cantidad de estados ocultos y cantidad de símbolos en las observaciones elegidos para el modelo, de esta manera se mejora la probabilidad a posteriori del modelo  $P(O|\lambda)$ . La tercera opción consiste en la interpolación de dos o más modelos generados a partir de una misma base de datos.

## **CAPITULO 4**

### **ADQUISICIÓN DE LAS SEÑALES DE VIBRACIÓN**

La base de datos consta de pruebas realizadas a los rodamientos de máquinas rotativas, en las cuales se hicieron pruebas a piezas normales y defectuosas.

Los experimentos y toma de datos provienen del centro de investigación Case Western Reserve University, los datos cuentan con la aprobación y respaldo de los siguientes centros de desarrollo tecnológico, Rockwell science center área de automatización, Office of naval Research (ONR) y la Federation of american scientists (FAS) área de desarrollo naval aéreo.

Los experimentos se realizaron en un motor de 2 caballos de fuerza, el fabricante del motor es Reliance Electric. La toma de datos fue realizada a diferentes distancias entre los rodamientos del motor y el micrófono de adquisición de datos.

A los rodamientos del motor se les sembraron fallas utilizando un mecanismo de electroerosión EDM (Electro-discharge machining). Las fallas van desde 0.007 pulgadas de diámetro hasta 0.040 pulgadas de diámetro, las cuales fueron introducidas en cada una de las componentes de los rodamientos por separado, el rodamiento está conformado por las siguientes partes: la pista de rodadura interior, la pista de rodadura exterior, y los elementos rodantes (balines). Estos elementos fueron reinstalados bajo algún nivel de falla a la hora de hacer las pruebas en el motor, el registro de vibraciones se hizo con el motor operando con cargas de 0 a 3 caballos de fuerza con velocidades en el motor de 1797rpm a 1720rpm.

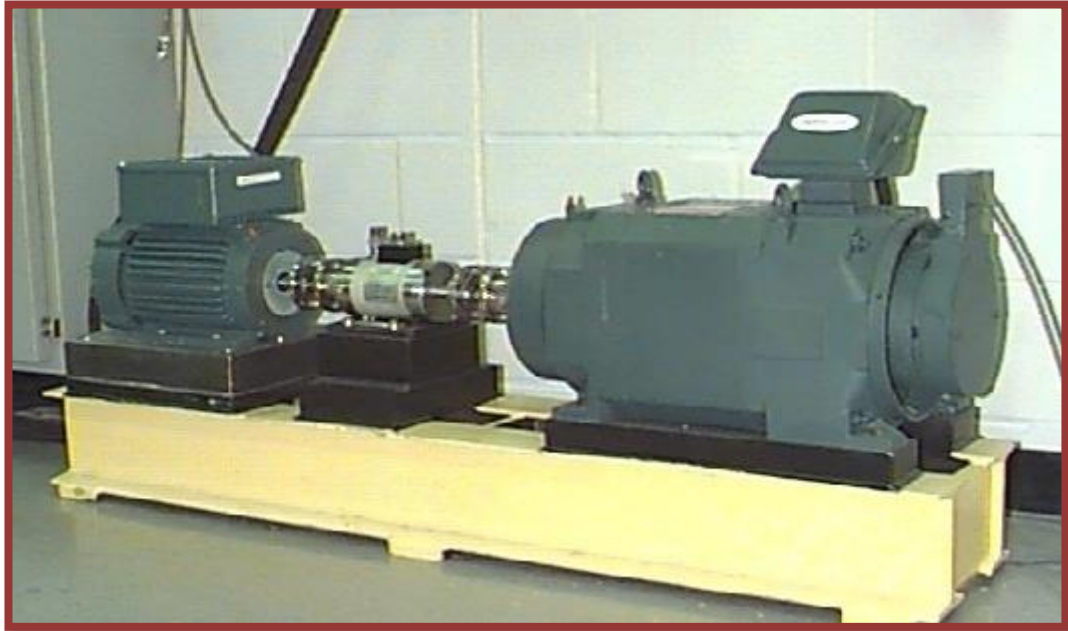
#### **Proceso descrito por el laboratorio para la adquisición de datos**

Los datos de vibraciones se recogieron mediante acelerómetros, los cuales fueron adheridos de forma magnética. Los acelerómetros se colocaron en la posición de las 12 horas, tanto en el lado de accionamiento del ventilador y al final de la carcasa del motor. Durante algunos experimentos, el acelerómetro se unió al soporte del motor. Las señales de vibración se recopilaban mediante un grabador DAT de 16 canales, y fueron después procesados en un entorno de Matlab. Todos los archivos de datos están en Matlab en formato (\*. Mat). Los datos digitales se recogieron a 12.000 muestras por segundo, este proceso se llevo a cabo para cada una de las especificaciones de falla. Los datos de velocidad y potencia fueron recolectados a través del transductor/codificador de par y se registraron manualmente [46].

Las fallas externas de rodadura fueron incluidas en la toma de datos, esto se hizo con el fin de asemejar el sistema lo mejor posible a un proceso industrial real, por lo cual la colocación de la falla con respecto a la zona de carga del rodamiento tiene un impacto directo sobre la respuesta del sistema. Con el fin de cuantificar este efecto, los experimentos se llevaron a cabo tanto para

la unidad del ventilador como para los cojinetes de rodadura, las fallas se presentaron en zonas externas ubicadas en 3 horas (directamente en la zona de carga), a las 6 horas (ortogonal a la zona de carga) y 12:00.

Figura 5. Sistema de pruebas a rodamientos.



### **Asignación de severidad por falla en rodamientos.**

#### **Severidad 1**

La severidad 1 (fisura de 0.007 pulgadas), representa desgaste en la sección del rodamiento al cual haga referencia, para este caso se considera que la máquina no se encuentra en falla y que su funcionamiento aún es aceptable, se puede entonces informar el tiempo restante para la máquina antes de un mantenimiento y también se puede dar una reseña de la tendencia que tiene la máquina de fallar en esta sección del rodamiento, esto se puede afirmar debido a que las fallas son degenerativas progresivas. El tiempo posible de falla se podría ajustar teniendo un historial del comportamiento entre fallas de la máquina, sin embargo para este trabajo de grado no se cuenta con histogramas de las máquinas de las cuales procede las señales de vibraciones por lo cual se omite este paso.

#### **Severidad 2**

La severidad 2 (fisura de 0.014 pulgadas), representa un desgaste serio en la sección del rodamiento, para este caso la máquina no se encuentra bajo riesgo de falla permanente pero su funcionamiento no es aceptable, con lo cual se indica que se debe realizar mantenimiento lo antes posible. Se debe informar que la máquina esta desgastada altamente en esta sección.

**Severidad 3**

La severidad 3 (fisura de 0.021-0.028 pulgadas) indica que la máquina se encuentra en una falla altamente riesgosa para la máquina aunque aún pueda funcionar, en este estado se debe bloquear la máquina e impedir su funcionamiento hasta que no se le haga una rutina de mantenimiento, esto se hace debido a que mantenerla en funcionamiento puede llevar a daños irreversibles en la máquina, y afectar la vida útil. También puede poner en riesgo el sistema dependiente de ella, en este estado se requiere el reemplazo total de la pieza, ya que el desgaste no puede ser remediado por las rutinas de mantenimiento realizables.



**PARTE III**  
**RESULTADOS Y CONCLUSIONES**

## **CAPITULO 5**

### **RESULTADOS**

En este capítulo se presenta la metodología desarrollada para el entrenamiento del sistema de detección y clasificación de fallas en los rodamientos de una máquina rotativa basada en los modelos ocultos de Markov con mezclas gaussianas (MHMM) y en los Procesos Gaussianos (GP). El entrenamiento del sistema de detección y clasificación de fallas se plantea a partir de una base de datos la cual se anota en cuatro posibles estados, los cuales son: falla en la cara externa del rodamiento, falla en la cara interna del rodamiento, falla en los balines del rodamiento y un estado de no falla o de funcionamiento aceptable. Adicional a esto para cada estado de falla se tiene 3 severidades. En el transcurso del capítulo se mencionaran los criterios de entrenamiento que se tuvieron en cuenta pensando en una implementación real.

#### **5.1 DISEÑO DE LA METODOLOGÍA PARA EL ENTRENAMIENTO DEL SISTEMA DE DETECCIÓN DE FALLAS**

##### **5.1.1 Metodología para la extracción de características relevantes**

El primer paso que se lleva a cabo para el entrenamiento del sistema de detección de fallas es procesar la base de datos de Bearing Data Center [46]. En la base de datos dispuesta se cuenta con datos de señales provenientes de tres secciones del motor bajo estudio, estas son: señales provenientes del ventilador del motor eléctrico, señales provenientes de la base del motor eléctrico y señales provenientes de la armadura del motor eléctrico, para realizar la tarea de entrenamiento, se utiliza las señales provenientes de la base del motor.

##### **Normalización de la base de datos**

El primer paso que se realiza es tomar la base de datos y normalizarla, esto se hace con el fin de evitar inconvenientes a futuro, como desbordamientos o entrenamientos inadecuados, con lo cual se mejora la calidad de los datos.

Al llevar a cabo el proceso de normalización, se cumple dos requisitos básicos que son:

- Evitar las pérdidas en la información.
- Evitar la pérdida de las dependencias funcionales.

### **Cambio de Escala de la frecuencia a la escala de Mel**

Como se presentó en el capítulo 4, la base de datos se encuentra en el dominio del tiempo, se hace una transformación al dominio de la frecuencia, la cual sigue no siendo ideal para llevar a cabo una tarea de entrenamiento, puesto que se puede arriesgar a perder características valiosas [5, 28] con el fin de evitar un entrenamiento erróneo, se hace/aplica un cambio de escala, que tiene como objetivo pasar los datos del dominio de la frecuencia a la escala de Mel, que es finalmente la escala ideal para estudiar patologías, señales sonoras u ondas cuasi-estacionarias [6, 14], siempre y cuando estas estén contenidas en el intervalo de frecuencia de 8 KHz a 12 KHz.

El procedimiento que se aplica a la base de datos es el descrito en el capítulo 2, una vez realizado este procedimiento se obtienen los MFCC del conjunto de señales de vibraciones del motor.

#### **5.1.2 Creación de los libros de códigos.**

Una vez extraídos los coeficientes Cepstrales, se puede decir que posee una base de datos apta para trabajar con el MHMM y el GP. Con esta nueva base de datos se crean los libros de códigos, como se describe en los Capítulos 2.3 y 2.4, en donde se hace uso de las técnicas Fuzzy K-Means y Spectral Clustering, las cuales fueron elegidas por la versatilidad que poseen para agrupar bases de datos complejas, como lo son el reconocimiento de voz y de imágenes.

Con la creación de los libros de códigos se pretende reducir el tiempo de ejecución de los algoritmos, como así mismo proveer un filtro más a la base de datos para alcanzar mejores resultados de confiabilidad al momento de presentar un reporte del estado de falla del rodamiento.

### **Validación de datos**

La técnica de validación de datos que se aplicó al entrenamiento es la descrita en 3.3.2, en donde se propone una técnica de validación que es la mezcla entre dos metodologías, la razón de esta mezcla es tener un conjunto de datos que garantice independencia en cada iteración, al conformarse los grupos de entrenamiento y de prueba, al mismo tiempo lograr que fuera lo suficientemente rigurosa para llevar una buena labor de entrenamiento de los modelos de clasificación minimizando el consumo de recursos computacionales, con la mezcla de estos dos procesos se garantiza estas condiciones porque se realiza una única iteración sobre los conjuntos, los cuales son el conjunto de entrenamiento y el conjunto de prueba.

En este proceso se decidió elegir el 70% de las muestras de las señales para el conjunto de entrenamiento y el 30% restante fueron las que se dejaron para conformar el conjunto de pruebas.

### **5.1.3 Entrenamiento de los Procesos Gaussianos y los HMM para el sistema de detección y clasificación**

En este paso se llevó a cabo la elección del tipo de HMM que se utiliza para entrenar el sistema de detección de fallas en los rodamientos del motor, el modelo que se eligió para entrenar el sistema de detección fue un HMM con mixturas gaussianas (MHMM).

Los procesos Gaussianos (GP) se entrenaron mediante un modelo basado en los ensayos de Bernoulli, en los cuales se tiene una verificación de pertenencia de cada variable con respecto a pares formados por el modelo al cual debe pertenecer, este modelo de entrenamiento se explica detalladamente en el capítulo 3.1.

Los criterios de evaluación que se proponen para optimizar los modelos son:

- El tiempo que toma el proceso de entrenamiento de los MHMM y los GP para el sistema de detección, este tiempo se mide en segundos, aunque se tiene en cuenta que este proceso no es el más adecuado, debido a que es un tiempo relativo con respecto a los parámetros del equipo, lo ideal es hacer un análisis de complejidad, pero este no se realiza en este trabajo.
- El segundo criterio de evaluación que se asumió, es maximizar la probabilidad de reconocimiento del sistema con respecto a la señal bajo estudio (maximizar la función de máxima verosimilitud) el cual se considera para este trabajo de maestría el enfoque verdadero de la investigación, ya que lo que se busca es tener resultados confiables del estado de desgaste del rodamiento.

Estos fueron los dos puntos clave que se plantearon, al llevar la labor de entrenamiento del sistema de detección de fallas en los rodamientos.

#### **Elección de las matrices de parámetros iniciales para los MHMMs.**

La elección de los parámetros iniciales para el sistema de detección (entrenamiento de los MHMM), se realizó de manera aleatoria generando un conjunto de variables aleatorias que surgen del valor de la media de las observaciones, así se genera la matriz de transición inicial  $Q$ , donde  $Q$  representa la cantidad de estados, de igual manera se generan los estados iniciales los cuales poseen dimensiones de  $Q \times 1$ , ambas matrices deben estar adecuadamente normalizadas, a partir de estas procedemos a generar las matrices de parámetros iniciales, en la cual adicionalmente se debe indicar el tipo de covarianza que se desea tener, entre estas

tenemos la covarianza completa (full), la diagonal (diag) y la covarianza esférica (spherical). En este cálculo se hace uso de la teoría de K-means clustering para estimar estos parámetros iniciales. El entrenamiento de los parámetros se llevó a cabo mediante el uso del algoritmo de Expectation Maximization (EM) descrito en el anexo B.

### **Elección de las matrices de parámetros iniciales para los Procesos Gaussianos.**

La elección de las matrices de parámetros iniciales para este caso se han nombrado hiper-parámetros, los cuales se pueden calcular mediante la aproximación por Laplace, este método es ampliamente descrito en el capítulo 3.1 en donde se hace referencia a las consideraciones que se deben tomar al momento de calcular los hiper-parámetros.

## **5.2 DISCUSIÓN ACERCA DEL SISTEMA DE DETECCIÓN Y CLASIFICACIÓN DE FALLAS**

Finalmente el sistema de detección y clasificador de fallas en los rodamientos de una máquina rotativa, consiste en el conjunto de MHMM y PG entrenados teniendo en cuenta todos los parámetros de máximo desempeño planteados anteriormente, para realizar la tarea de detección y clasificación de fallas, se evaluó a que estado pertenece entre los diez planteados una señal captada de una máquina rotativa semejante a la cual se realizó el análisis, los posibles estados de falla se muestran a continuación, estos mismos fueron expuestos anteriormente cuando se planteó que el sistema podría ser más robusto permitiéndonos llevar a cabo una labor más compleja como lo es la clasificación de fallas.

- I. Estado de no falla
- II. Falla en los balines del rodamiento severidad 1 (fisura de 0.007 pulgadas)
- III. Falla en los balines del rodamiento severidad 2 (fisura de 0.014 pulgadas)
- IV. Falla en los balines del rodamiento severidad 3 (fisura de 0.021-0.028 pulgadas)
- V. Falla en la cara interna del rodamiento severidad 1 (fisura de 0.007 pulgadas)
- VI. Falla en la cara interna del rodamiento severidad 2 (fisura de 0.014 pulgadas)
- VII. Falla en la cara interna del rodamiento severidad 3 (fisura de 0.021-0.028 pulgadas)
- VIII. Falla en la cara externa del rodamiento severidad 1 (fisura de 0.007 pulgadas)
- IX. Falla en la cara externa del rodamiento severidad 2 (fisura de 0.014 pulgadas)
- X. Falla en la cara externa del rodamiento severidad 3 (fisura de 0.021-0.028 pulgadas)

Hasta el momento no se había dado una razón clara de porque el sistema podría desempeñar adicionalmente a su labor de detección, una de clasificación de fallas, la razón es simple, al poseer más estados de fallas, se puede inferir mediante un historial de labores de mantenimiento y la captación de nuevas señales, cual es la falla actual en la máquina rotativa. Así las severidades indicaran la evolución de la máquina.

### 5.2.1 Resultados obtenidos para los MHMM.

A continuación se presentan los resultados obtenidos al aplicar el procedimiento descrito anteriormente al usar los MHMM sobre la base de datos generada como producto del tratamiento de los datos.

Para cada caso se repitió el experimento 10 veces, de los cuales se calculó el promedio de los resultados. El conjunto de entrenamiento y validación fue elegido de manera diferente y aleatoria, impartiendo así el método de validación híbrido diseñado en este mismo proyecto, el cual es explicado en el capítulo 3.

Es importante tener entonces presente que cada una de las tablas que se presentan a continuación, poseen filas y columnas en las cuales se encuentran mencionados los diez posibles estados de desgaste en los cuales se puede encontrar el rodamiento. Sin embargo para estas tablas las filas representan la procedencia de las señales entrantes a las pruebas en los modelos. Las columnas representan el modelo con el cual se llevó a cabo la prueba de pertenencia, estas tablas es lo que se conocen como tablas de confusión.

El sistema de clasificación es entrenado usando tres diferentes tipos de bases de datos, en estas bases de datos tenemos el libro de código generado a partir de la técnica de Fuzzy K-Means, el libro de código generado a partir de Spectral Clustering y finalmente tenemos la base de datos sin discretizar.

Para la base de datos generada a partir de Fuzzy K-Means se usaron 10 núcleos o 10 estados, ahora para la creación del libro de códigos creado a partir de Spectral Clustering se asignaron 7 núcleos.

La cantidad de mezclas Gaussianas utilizada para la generación del modelo se fijó en  $M=4$  y la cantidad de estados ocultos para este mismo fue de 4,  $Q=4$ .

### Resultados con los datos del Fuzzy K-Means sobre el clasificador MHMM

Tabla 1. Resultados con los datos del Fuzzy K-Means sobre el clasificador MHMM

Category		No failure	Ball			Inner ring			Outer ring		
			1	2	3	1	2	3	1	2	3
No failure		96% ± 4%	2% ±11%	0%	0%	0%	0%	0%	2% ±11%	0%	0%
Ball	1	0%	96% ±4%	2% ±10%	2% ±10%	0%	0%	0%	0%	0%	0%
	2	0%	2% ±10%	96% ± 4%	2% ±10%	0%	0%	0%	0%	0%	0%
	3	0%	2% ±8%	1% ±8%	97% ±3%	0%	0%	0%	0%	0%	0%

Inner ring	1	0%	0%	0%	0%	96% ±4%	2% ±10%	2% ±8%	0%	0%	0%
	2	0%	0%	0%	0%	%	100% ± 0%	0%	0%	0%	0%
	3	0%	0%	0%	0%	0%	0%	99% ± 1%	0%	0%	0%
Outer ring	1	0%	0%	0%	0%	0%	0%	0%	100% ± 0%	0%	0%
	2	0%	0%	0%	0%	0%	0%	0%	0%	100% ± 0%	0%
	3	0%	0%	0%	0%	0%	0%	0%	0%	0%	100% ± 0%

En la tabla 1 se puede observar los resultados obtenidos después de realizar diez entrenamientos del modelo para el MHMM usando la base de datos difusa, al observar los resultados este modelo es un  $96\% \pm 4\%$  efectivo reconociendo el estado de falla.

### Resultados con los datos originales sobre el clasificador MHMM

En la tabla 2 se puede observar los resultados obtenidos después de realizar diez entrenamientos del modelo MHMM usando la base de datos original, la base de datos original realmente hace referencia a los datos a los cuales se les aplico el cambio de dominio del tiempo a el dominio de Mel Cepstrum, el cual es una sub-escala de la frecuencia, en donde los posibles valores se ubican en la escala de frecuencia auditiva humana. El clasificador se equivoca en un  $7\% \pm 11\%$  al clasificar, para este modelo el porcentaje de reconocimiento de las señales está en un  $93\% \pm 7\%$  de efectividad, el cual es menos confiable que para el caso anterior.

Tabla 2. Resultados con los datos originales sobre el clasificador MHMM

Category		No failure	Ball			Inner ring			Outer ring		
			1	2	3	1	2	3	1	2	3
No failure		100% ± 0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ball	1	0%	93% ±7%	7% ±11%	0%	0%	0%	0%	0%	0%	0%
	2	0%	0%	100% ± 0%	0%	0%	0%	0%	0%	0%	0%
	3	0%	0%	0%	100% ± 0%	0%	0%	0%	0%	0%	0%
Inner ring	1	0%	0%	0%	0%	100% ± 0%	0%	0%	0%	0%	0%
	2	0%	0%	0%	0%	0%	100% ± 0%	0%	0%	0%	0%
	3	0%	0%	0%	0%	0%	0%	100% ± 0%	0%	0%	0%
Outer ring	1	0%	0%	0%	0%	0%	0%	0%	100% ± 0%	0%	0%

	2	0%	0%	0%	0%	0%	0%	0%	0%	100% ± 0%	0%
	3	0%	0%	0%	0%	0%	0%	0%	0%	0%	100% ± 0%

## Resultados con los datos del Spectral Clustering sobre el clasificador MHMM

Tabla 3. Resultados con los datos del Spectral Clustering sobre el clasificador MHMM

Category		No failure	Ball			Inner ring			Outer ring		
			1	2	3	1	2	3	1	2	3
No failure		0%	0%	45% ±11%	0%	0%	10% ±22%	20% ±11%	15% ±14%	5% ±11%	5% ±11%
Ball	1	0%	0%	20% ±11%	0%	0%	5% ±11%	0%	15% ±14%	50% ±0%	10% ±14%
	2	0%	0%	25% ±18%	0%	0%	5% ±11%	20% ±11%	0%	50% ±0%	0%
	3	0%	0%	35% ±22%	0%	10% ±22%	0%	30% ±27%	0%	25% ±0%	0%
Inner ring	1	20% ±11%	0%	0%	0%	40% ±14%	5% ±11%	0%	15% ±14%	10% ±14%	10% ±14%
	2	0%	0%	35% ±22%	0%	0%	20% ±21%	0%	0%	40% ±14%	0%
	3	20% ±11%	0%	0%	0%	27% ±4%	0%	25% ±18%	15% ±14%	10% ±14%	5% ±11%
Outer ring	1	0%	0%	0%	5%	0%	0%	40% ±22%	40% ±42%	5% ±11%	10% ±14%
	2	0%	0%	0%	0%	0%	0%	10% ±14%	15% ±14%	75% ±0%	0%
	3	0%	0%	5% ±11%	0%	10% ±14%	25% ±18%	0%	0%	5% ±11%	55% ±21%

En la tabla 3 se presentan los resultados del experimento final con respecto a los MHMM, al igual que en las anteriores pruebas se realizó diez diferentes entrenamientos y en la tabla 3 se condensan los resultados conjuntos de estos, como se puede observar al usar la base de datos tratada mediante los Spectral Clustering en el cual se tiene un valor de porcentaje de clasificación general alrededor del 55% ± 21%, debido a esto, este método debe ser descartado para este tipo de labor, ya que los valores de clasificación que se obtienen no son factibles.

### 5.2.2 Resultados obtenidos para los GP.

A continuación se discutirán los resultados obtenidos al aplicar el procedimiento descrito anteriormente al usar los GP sobre las tres bases de datos generadas como producto del tratamiento de los datos originales.



Para cada caso se repitió el experimento 10 veces, de los cuales se calculó el promedio de los resultados. El conjunto de entrenamiento y validación fue elegido de manera diferente y aleatoria, impartiendo así el método de validación híbrido diseñado en este mismo proyecto, el cual es explicado en el capítulo 3.

Es importante tener entonces presente que cada una de las tablas que se presentan a continuación, poseen filas y columnas en las cuales encontramos mencionados los diez posibles estados de desgaste en los cuales se puede encontrar el rodamiento. Sin embargo para estas tablas las filas representan la procedencia de las señales entrantes a las pruebas en los modelos, por lo cual las columnas representan el modelo con el cual se hace la prueba de pertenencia, también es importante recordar que estas tablas se hicieron en base a un modelo binario, en donde la señal solo puede estar en dos opciones las cuales son: pertenecer al grupo de la cual procede o pertenecer al grupo incorrecto en cuestión.

### Resultados con los datos del Fuzzy K-Means sobre el clasificador GP

Tabla 4. Resultados con los datos del Fuzzy K-Means sobre el clasificador GP

Category		No failure	Ball			Inner ring			Outer ring		
			1	2	3	1	2	3	1	2	3
No failure		90% ±10%	0%	0%	5% ±14%	2.5% ±11%	2.5% ±11%	0%	0%	0%	0%
Ball	1	0%	95% ±5%	2.5% ±11%	0%	0%	0%	0%	0%	2.5% ±11%	0%
	2	2% ±21%	1% ±14%	77% ±11%	3% ±11%	8% ±21%	1% ±21%	0%	0%	8% ±33%	0%
	3	6% ±18%	6% ±11%	0%	78% ±11%	5% ±11%	2% ±27%	1% ±14%	6% ±25%	1% ±14%	0%
Inner ring	1	2.5% ±14%	0%	0%	0%	95% ±5%	0%	2.5% ±11%	0%	0%	0%
	2	4% ±0%	4% ±18%	0%	0%	0%	77% ±11%	2% ±21%	5% ±11%	8% ±11%	0%
	3	0%	0%	0%	0%	0%	0%	98% ±2%	2% ±14%	0%	0%
Outer ring	1	0%	0%	0%	10% ±21%	0%	0%	6% ±14%	84% ±11%	0%	0%
	2	2.5% ±11%	2.5% ±11%	0%	0%	0%	5% ±14%	0%	0%	90% ±10%	0%
	3	0%	0%	3% ±11%	0%	10% ±21%	0%	0%	0%	3% ±11%	84% ±11%

En la tabla 4 se tiene los resultados obtenidos tras hacer las pruebas con los datos tratados mediante la discretización por los Fuzzy K-Means, en estos resultados se puede observar la diagonal que posee el valor de clasificación que se necesita, esta representa el valor de reconocimiento de las señales al ser pasadas por el modelo entrenado mediante ese tipo, ahora mirando nuevamente la fila 1, se puede observar valores del 5% ± 14% de reconocimiento del

compendio de señales en ball severidad 3 al igual que un  $2.5\% \pm 11\%$  de reconocimiento de las señales en Inner ring para severidades 1 y 2.

Volviendo nuevamente al análisis de la tabla 4, se puede decir que el valor de reconocimiento para este modelo planteado se encuentra en un  $77\% \pm 11\%$  de confiabilidad de detectar correctamente el estado de la maquina bajo los parámetros mencionados anteriormente.

## Resultados con los datos originales sobre el clasificador GP

Tabla 5. Resultados con los datos originales sobre el clasificador GP

Category		No failure	Ball			Inner ring			Outer ring		
			1	2	3	1	2	3	1	2	3
No failure		100% $\pm 0\%$	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ball	1	0%	100% $\pm 0\%$	0%	0%	0%	0%	0%	0%	0%	0%
	2	0%	3% $\pm 11\%$	97% $\pm 3\%$	0%	0%	0%	0%	0%	0%	0%
	3	0%	0%	0%	100% $\pm 0\%$	0%	0%	0%	0%	0%	0%
Inner ring	1	0%	0%	0%	0%	100% $\pm 0\%$	0%	0%	0%	0%	0%
	2	0%	0%	0%	0%	0%	100% $\pm 0\%$	0%	0%	0%	0%
	3	0%	0%	0%	0%	0%	0%	100% $\pm 0\%$	0%	0%	0%
Outer ring	1	0%	0%	0%	0%	0%	0%	0%	100% $\pm 0\%$	0%	0%
	2	0%	0%	0%	0%	0%	0%	0%	0%	100% $\pm 0\%$	0%
	3	0%	0%	0%	0%	0%	0%	0%	0%	0%	100% $\pm 0\%$

En la tabla 5 se presentan los resultados obtenidos al presentar las señales originales al clasificador Gaussiano, para el cual se obtuvieron resultados de confiabilidad del  $97\% \pm 3\%$  en la tarea de reconocimiento. Por lo tanto, es posible elaborar una rutina de mantenimiento.

A diferencia de los MHMM que son más eficientes en la tarea de reconocimiento, los PG son de rápido entrenamiento, por ser computacionalmente más eficientes al estar basados en el ensayo de Bernoulli, por lo cual si se desea aplicar una rutina para hacer pruebas en caliente, o diagnóstico y pronóstico online, para lo cual este método sería el indicado.

## Resultados con los datos del Spectral Clustering sobre el clasificador GP

Tabla 6. Resultados con los datos del Spectral Clustering sobre el clasificador GP

Category		No failure	Ball			Inner ring			Outer ring		
			1	2	3	1	2	3	1	2	3
No failure		62% ±11%	3% ±18%	3% ±18%	10% ±18%	1% ±18%	3% ±18%	10% ±53%	3% ±18%	3% ±18%	2% ±0%
Ball	1	3% ±18%	62% 12%	3% ±18%	10% ±18%	3% ±0%	3% ±0%	3% ±18%	3% ±18%	10% ±0%	0%
	2	3% ±18%	4% ±18%	72% ±11%	3% ±18%	0%	3% ±18%	0%	6% ±53%	3% ±18%	6% ±35%
	3	0%	3% ±18%	10% ±35%	66% ±14%	4% ±35%	8% ±53%	3% ±18%	0%	6% ±0%	0%
Inner ring	1	10% ±18%	5% ±35%	2% ±0%	2% ±35%	62% ±21%	2% ±35%	5% ±35%	2% ±35%	4% ±53%	2% ±18%
	2	3% ±18%	15% ±0%	1% ±18%	15% ±35%	1% ±18%	57% ±21%	2% ±18%	2% ±18%	2% ±18%	2%
	3	2% ±18%	2% ±18%	2% ±0%	6% ±18%	2% ±18%	2% ±18%	80% ±11%	2% ±18%	2% ±35%	0%
Outer ring	1	3% ±53%	3% ±35%	3% ±35%	3% ±18%	3% ±18%	10% ±0%	3% ±18%	66% ±21%	3% ±35%	3% ±0%
	2	0%	4% ±0%	0%	3% ±18%	4% ±0%	3% ±18%	0%	3% ±18%	80% ±11%	3% ±35%
	3	7% ±35%	3% ±53%	7% ±0%	3% ±18%	3% ±18%	3% ±53%	3% ±53%	3% ±0%	3% ±53%	65% ±12%

Finalmente se presenta la tabla 6, la cual posee los resultados del último experimento realizado. En este caso la labor de clasificación es totalmente inaceptable, dado que la clasificación da patrones desordenados y aleatorios, lo cual demuestra que al aplicar los Spectral Clustering tanto en los GP como en los MHMM el clasificador reconoce el 50% de manera correcta, por lo cual la tarea de clasificación no se lleva a cabo.

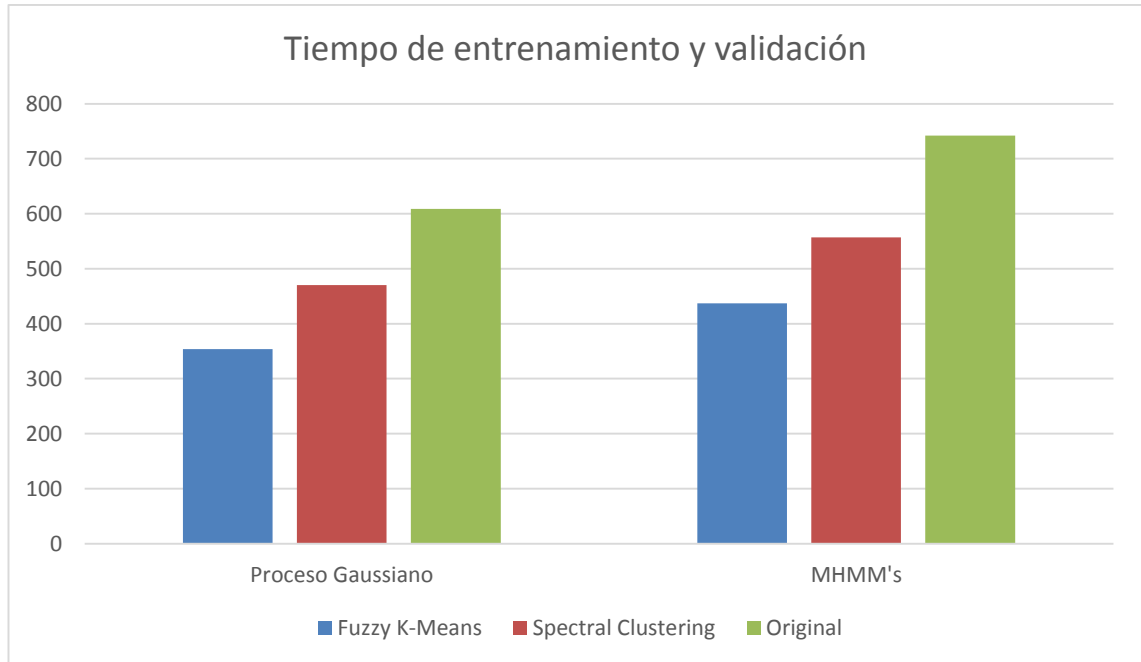
### 5.2.3 Discusión de resultados

Anteriormente se presentaron las tablas con los resultados de las pruebas realizadas al variar tanto las señales de prueba (Code Books) como al cambiar entre los dos diferentes modelos usados para entrenar los clasificadores.

De estos resultados se puede concluir, que el mejor clasificador para realizar el pronósticos de vida útil de los rodamientos por confiabilidad es el MHMM al usar las señales del Code Book creado a partir del método de los Fuzzy K-Means. Sin embargo por otra parte, cuando hablamos de eficiencia con respecto al tiempo tomado para obtener un resultado de clasificación, teniendo en cuenta una buena tasa de confiabilidad, debemos decir que el modelo que mejor se adapta para esta labor es el GP al hacer uso de las señales originales (las que solo poseen la

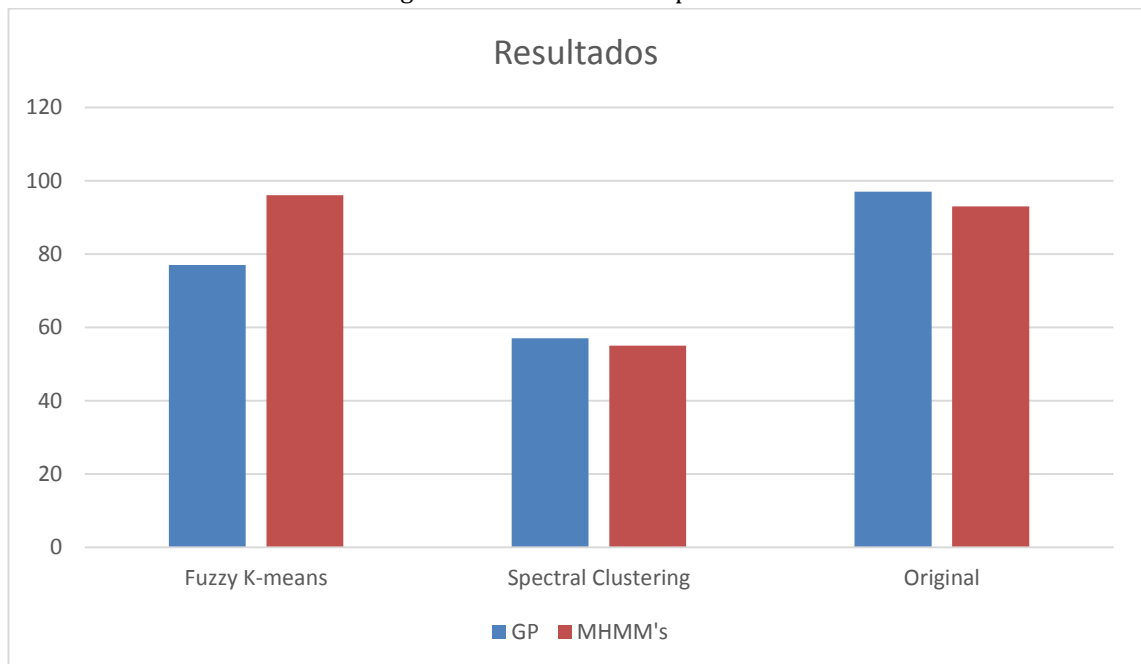
extracción de características por Mel Cepstrum), ya que no requiere un filtrado adicional, como así mismo podemos asegurar que el entrenamiento de un modelo basado en un GP es de menor complejidad que un MHMM. El valor del tiempo está dado en segundos y se presenta en la figura 6.

Figura 6. Tiempo de entrenamiento y validación.



A continuación se presenta un diagrama de barras, el cual muestra en resumen los resultados de la investigación, mostrando la ventaja y desventaja de una técnica con respecto a otra.

Figura 7. Resultados de la práctica.



Después de observar el gráfico, nuevamente cabe destacar que los peores casos se presentan al hacer uso de los datos discretizados mediante los Spectral Clustering y que en general los mejores resultados dan al tratar la base de datos como originalmente quedan después de usar el filtrado por Mel.

Finalmente como aporte al grupo de investigación, en el área de mantenimiento se ha desarrollado una técnica para realizar el diagnostico de un rodamiento con desgaste, tanto para aplicaciones en caliente como para aplicaciones que se puede realizar al final de una jornada laboral. Mediante estos resultados podemos entonces plantear las técnicas de pronóstico también, ya que conociendo el estado actual de desgaste, más un historial de desgaste del rodamiento se puede fácilmente calcular los tiempos para realizar las prácticas de mantenimiento, sin interferir con la producción por fallas presentadas sin aviso, cuando las operaciones no se salen de las prácticas habituales.

### **5.3 PLANTEAMIENTO DEL SISTEMA DE DIAGNOSTICO Y PRONOSTICO DE FALLAS CON BASE EN LOS RESULTADOS OBTENIDOS POR LOS CLASIFICADORES.**

Finalmente al tener los resultados de los clasificadores, podemos plantear las rutinas de mantenimiento en base al RUL. El RUL se calcula considerando el estado de desgaste actual del rodamiento y haciendo uso de los datos entregados por el fabricante, este método aplica para cualquier bien material que se posea, que pueden ir desde rodamientos hasta sistemas industriales complejos.

Los rodamientos usados para generar la base de datos son del fabricante SKF[6], por lo cual se debe recurrir a ellos y solicitar las bases de datos correspondientes a los tiempos de vida útil de los rodamientos, estos datos generalmente poseen historiales completos del funcionamiento de todos los rodamientos bajo diferentes tipos de cargas laborales, así mismo como de las condiciones ambientales a las cuales son sometidos, como por ejemplo, ambientes con alta humedad o altas temperaturas, por lo cual el ingeniero de mantenimiento debe tomar en cuenta estos parámetros al momento de acceder a consultar estos datos.

El sistema de diagnóstico se plantea llevando una comparación del estado de desgaste actual del rodamiento dados los resultados del clasificador, el tipo de carga laboral que lleve la máquina y el ambiente de trabajo en el cual se encuentre, así teniendo en cuenta estos tres aspectos el ingeniero de mantenimiento puede calcular el RUL del rodamiento.

Con el cálculo del RUL que se hace mediante el diagnóstico, se puede preparar el pronóstico del rodamiento, el cual tiene por objetivo optimizar las rutinas de mantenimiento y prolongar en la medida que sea posible la utilidad del rodamiento bajo los parámetros de calidad y seguridad que permita la labor desempeñada por él, que generalmente consisten en el tipo de proceso industrial al cual pertenezca, así como en qué tan crítico es que se presente una falla por desgaste no pronosticada.

## **CAPITULO 6**

### **CONCLUSIONES Y TRABAJOS FUTUROS**

- Al implementarse los Code Books sobre las bases de datos después de aplicar las técnicas de extracción de características en señales cuasi-estacionarias, como lo es la técnica del cambio de escala de la frecuencia a la escala de Mel, para la extracción de coeficientes cepstrales (MFCC), se da una mejora en tiempo de cálculo de los sistemas de clasificación, sin embargo adicional a esto en el caso del MHMM con datos difusos también se presentó un mejoramiento en la confiabilidad de la clasificación ya que se venía trabajando con un modelo que estaba arrojando un valor de confiabilidad en la clasificación del 95% con los datos no tratados hasta el 96%, con lo cual se obtuvieron resultados mejores de lo esperado.
- La metodología de detección y clasificación de fallas diseñada en base a los MHMM y los GP realizada en este proyecto de grado, permite con base en una población muestral de datos anotada generar modelos de inferencia para estimar los parámetros que modelan las fallas en un sistema bajo estudio, esta metodología asegura buenos resultados en los porcentajes de detección y clasificación en el sistema, siempre y cuando las bases de datos captadas sean de procedencia cuasi-estacionaria. Igualmente se atribuye al sistema diseñado un buen desempeño en aplicaciones reales debido al poco tiempo y recursos computacionales requeridos en las tareas de detección y clasificación sobre todo cuando trabajamos con los Code Books en el clasificador diseñado mediante los Procesos Gaussianos.
- Hallar los parámetros necesarios para la implementación de los modelos en la tarea de identificación y clasificación de fallas se hizo de forma concisa debido a los criterios de optimización impuestos en la metodología, lo cual demuestra que estos fueron razonables tanto por la parte de optimización de recursos computacionales como en la confiabilidad de los modelos inferidos.
- La implementación de estas metodologías de mantenimiento permiten mejorar las condiciones de vida en una máquina o sistema, debido a que se está previniendo el desgaste temprano de una máquina o sistema al realizarse de manera oportuna las prácticas de mantenimiento, lo cual disminuye considerablemente la intervención humana y el cambio de piezas temprano, esto demuestra el impacto positivo que se

tiene ya que se disminuye el costo por labores de mantenimiento, al mismo tiempo que se disminuye el impacto negativo al medio ambiente.

- En la tabla 1 se puede observar los resultados obtenidos después de realizar diez entrenamientos del modelo para el MHMM usando la base de datos difusa, al observar los resultados este modelo es un 97% efectivo reconociendo el estado de falla. Con lo cual se puede asegurar que al llevar a cabo un reporte de estado de falla después de obtener las señales y aplicar este clasificador nos entregara un resultado correcto de reconocimiento.

Algunos trabajos futuros que se pueden recomendar y que amplían el campo de aplicación de este trabajo de grado son:

- La implementación del sistema entrenado con base en MHMM para el caso de este proyecto de grado puede ampliarse a una aplicación industrial, una manera que se plantea para la ampliación de esta implementación es obteniendo muestras de señales de vibraciones de otras máquinas, y usarlas en el entrenamiento de los modelos, lo cual es permitido debido a que el sistema no presenta sobre-entrenamiento. Una técnica similar se implemente durante mi estadía en el grupo de investigación de análisis de vibraciones de la facultad de mecánica.
- Usar el análisis de componentes principales u otras técnicas afines para mejorar la calidad de las señales obtenidas después de aplicar un filtro o una técnica de discretización, para obtener resultados más confiables o poder tener una metodología más eficiente computacionalmente.



## BIBLIOGRAFÍA

- [1] Omid Geramifarda, Jian-Xin Xua, Sanjib Kumar Panda, Fault detection and diagnosis in synchronous motors using hidden Markov model-based semi-nonparametric approach, *Engineering Applications of Artificial Intelligence*, Vol. 26, Issue 8, September 2013, pp 1919–1929.
- [2] George Georgoulas, Theodore Loutas, Chrysostomos D. Stylios, Vassilis Kostopoulos, Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition, *Mechanical Systems and Signal Processing*, Vol. 41, Issues 1–2, May., 2013, pp. 510–525.
- Abu-Rukah, Y., Al-Kofahi, O., Assessment of the effect of landfill leachate on groundwater quality-a case study. El-Akader Landfill Site-North Jordan., *Journal of Arid Environments*, Vol. 49, No. 3, Nov., 2001, pp. 615-630.
- [3] Tony Boutros, Ming Liang, Detection and diagnosis of bearing and cutting tool faults using hidden Markov models, *Mechanical Systems and Signal Processing*, Vol. 25, Issue 6, August 2011, pp. 2102–2124.
- [4] K. A. Loparo, Seeded Fault Test Data, Bearing Data Center, Case Western Reserve University, <http://csegroups.case.edu/bearingdatacenter/pages/12k-drive-end-bearing-fault-data>
- [5] Kanungo T. Mount, D. M.; Netanyahu, N. S. Piatko, C. D. Silverman, R. Wu, A. Y. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Analysis and development of intelligence in machines*, Vol. 24, No. 2, July., 2002, pp 881–892.
- [6] <http://csegroups.case.edu/bearingdatacenter/pages/apparatus-procedures> consultada 10 de Noviembre de 2015.
- [7] C. E. Rasmussen & C. K. I. Williams, *Processes for Machine Learning*, MIT Press., 2006
- [8] R. Carl, N. Hannes, *The GPML Toolbox*, Version 3.2., 2013.
- [9] Diego Muñoz T., Diseño de una metodología para la detección de fallas en rodamientos de máquinas rotativas mediante el uso de Modelos Ocultos De Markov, Universidad Tecnológica, Pereira, 2012.
- [10] Xing-Hui Zhang., Jian-She Kang., Hidden Markov models in bearing fault diagnosis and prognosis., *Computational Intelligence and Natural Computing Proceedings (CINC)*, IEEE conference publications Vol 2., 2010 , pp 364-367
- [11] <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file> consultada el 10 de Noviembre del 2015.

- [12] Ocak, H. ; Loparo, K.A., A new bearing fault detection and diagnosis scheme based on hidden Markov modeling of vibration signals., Acoustics, Speech, and Signal Processing, IEEE International Conference., Vol 5., 2001., pp 3141-3144
- [13] White D., Helicopter Usage Monitoring Using the MaxLife System. DSTO Workshop on Helicopter Health and Usage Monitoring System., Melbourne, Australia, February., 1999.
- [14] L. S. Chee, O. C. Ai, M. Hariharan, and S Yaacob., MFCC based recognition of repetitions and prolongations in stuttered speech using K-NN and LDA., in IEEE Student conference on Research and development., UPM Serdang., Malaysia., 2009., pp 46-149.
- [15] Boškoski, P.a, Gašperin, M.a, Petelin, D., Bearing fault prognostics based on signal complexity and Gaussian process models., IEEE International Conference on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application., United States., June., 2012., pp 1-8
- [16] Castellanos D Germán, Yuri S. Shinakov., Análisis De Aleatoriedad En Señales Y Sistemas., Automática avanzada., Manizales: Centro de Publicaciones UN Manizales, 2007.
- [17] <http://www.cse.unsw.edu.au/~waleed/phd/html/node38.html>, descargado el 3 de marzo del 2010.
- [19] Rabiner L., A Tutorial On Hidden Markov Models And Selected Applications In Speech Recognition., IEEE, Vol. 77, No 2, 1989., pp 257-285.
- [20] Lloyd, S. P. Least square quantization in PCM. Bell Telephone Laboratories Paper. Published in journal much later: Lloyd, S. P. Least squares quantization in PCM. IEEE Transactions on Information Theory, año 1995 páginas 129–137.
- [21] MacKay, David, Capítulo 20. An Example Inference Task: Clustering. Information Theory, Inference and Learning Algorithms. Universidad de Cambridge. Año 2003 página. 284–292.
- [22] Hamerly, G. y Elkan, C. Alternatives To The K-Means Algorithm That Find Better Clusterings. Procesos de la onceava conferencia internacional sobre conocimiento y manipulación de la información, procedimientos estadísticos, Año 2002.
- [23] Leonid B koralov y Yakov g. Sinai, Theory of Probability and Random Processes. Segunda edición, usa, año 2007.
- [24] S. P. Meyn y R.L. Tweedie. Markov Chains and Stochastic Stability. Segunda edición, universidad de Cambridge, año 2008.
- [25] Álvarez López Mauricio Alexander. Reconocimiento De Voz Sobre Diccionarios Reducidos Usando Modelos Ocultos de Markov, [Tesis para optar por el título de ingeniero electrónico].

Manizales, Universidad Nacional de Colombia Sede Manizales, Facultad de Ingeniería y Arquitectura, 2004, 71 páginas.

[27] Baum L. E. An Inequality And Associated Maximization Technique In Statistical Estimation For Probabilistic Functions Of A Markov Process, in Inequalities, volumen III, Editorial New York: Academic, año 1972, páginas. 1-8.

[28] Baum L. E.; Petrie T.; Soules G. y Weiss N. A Maximization Technique Occurring In The Statistical Analysis Of Probabilistic Functions Of Markov Chains, Volumen 41, año 1970, páginas. 164-171.

[29] Nadas A. Hidden Markov Chains, The Forward-Backward Algorithm And Initial Statistics, IEEE Truns. acústica. Procesamiento de señales del habla, volumen 3, Abril 1983, páginas 504-506.

[30] Viterbi A. J. Error Bounds For Convolutional Codes And An Asymptotically Optimal Decoding Algorithm, IEEE trnas. Teoría informativa, volumen 13, abril 1967, páginas 260-269.

[31] Forney G. D. The Viterbi algorithm, Proc IEEE, volumen 61, marzo del año 1973, páginas 228-278.

[32] Dempster A. P.; Laird N. M. y Rubin D. B. Maximum Likelihood From Imcomplete Data Via The EM Algorithm, volumen 39, año 1977, páginas 1-38.

[33] B. H Juan; S. E. Levinson, y M. M. Sondhi, Maximum Likelihood Estimation For Multivariable Mixture Observations Of Markov Chains, IEEE Trans. Volumen 32, no. 2, marzo del año de 1986, páginas 307-309.

[34] B. H Juan. Maximum Likelihood Estimation For Mixture Multivariate Stochastic Observations Of Markov Chains, volumen 64, no.6, julio año 1985, páginas 1235-1249.

[35] SP. Meyn y R.L Tweedi. Markov Chains and Stochastic Stability, Springer- verlag, Septiembre del año 2005.

[36] Salas Gonzáles Diego. Modelos De Mezclas De Distribuciones  $\alpha$ -Estables Aplicación A Micromatriz De Expresión Genética, [Tesis para optar al grado de doctor], España, Universidad de Granada, Departamento de física aplicada, Junio 2008, páginas 29-35.

[37] Gauvain Jean-Luc y Lee Chin-Hui. Bayesian Learning of Gaussian Mixture Densities for Hidden Markov Models, IEEE, departamento de investigación del habla, Bell Laboratories Murray Hill, año 2009

- [38] Carvajal González Johanna Paola. Metodología De Entrenamiento De Modelos De Mezclas Gaussianas Empleando Criterios De Gran Margen Para La Detección De Patologías En Bioseñales, [Tesis para optar al título de maestría], Manizales Colombia, Universidad Nacional de Colombia, Departamento de Ingeniería Eléctrica, Electrónica y Computación, Enero 2010.
- [39] Diulio M, Savage C, Finley B y Schneider E. Taking The Integrated Condition Assessment System To The Year 2010, Orlando, Florida año 2003.
- [40] Swamidas A y Chen Y. Damage Detection in a Tripod Tower Platform (TTP) Using Modal Analysis. ASME Offshore Technology, año 1992, páginas 577 a 583.
- [41] Bunks C; McCarthy D. y Al-Ani T. Condition-Based Maintenance of Machines Using Hidden Markov Models. Edición 14(4), año 2000 páginas 597 a 612.
- [42] Larsen Richard J. y Marx Morris L., An Introduction to Mathematical Statistics and Its Applications, tercera edición, año 2000 páginas 282.
- [43] Kittler J. Pattern Recognition: A Statistical Approach, Prentice-Hall, Londres, año 1982.
- [44] Geisser, Seymour. Predictive Inference. New York, editorial Chapman and Hall, año 1993
- [45] Kohavi Ron. A Study Of Cross-Validation And Boot Strap For Accuracy Estimation And Model Selection. Catorceava conferencia internacional sobre inteligencia artificial, Morgan Kaufmann, San Mateo, California, año 1995

## APÉNDICE A

### DESCRIPCIÓN MATEMÁTICA DEL ALGORITMO DE ENTRENAMIENTO

Dado un conjunto de observaciones  $\mathbf{O} = (\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_n)$  donde cada observación es una variable vectorial real de dimensión  $d$ , el algoritmo de *K-mean clustering* tiene como objetivo la partición de las  $n$  observaciones en  $k$  agrupaciones  $\mathbf{S} = (S_1, S_2, \dots, S_k)$  donde  $(k \leq n)$ . El objetivo del algoritmo es minimizar la cantidad de variables presentes en el espacio muestral [21, 22], el algoritmo está compuesto de 5 pasos computacionales que se explican a continuación:

#### Paso 1: Condiciones de inicio

Para la elección de las agrupaciones iniciales, se puede escoger entre dos metodologías, estas corresponden al método de particiones aleatorias y al método de Forgy [23]. El método de Forgy elige aleatoriamente  $k$  observaciones del conjunto de datos, con las cuales genera las agrupaciones iniciales. El método de particiones aleatorias asigna aleatoriamente un dato a cada agrupación, con lo que luego avanza al paso de actualización.

El método de Forgy tiende a divergir del conjunto de datos, mientras que la metodología de las particiones aleatorias tiende a converger a ellos. De acuerdo con Hammerly [23], el método de particiones aleatorias es preferible, en general.

#### Paso 2: Asignación

Asignar cada observación a la agrupación que asegure la mayor media.

$$J = \sum_{j=1}^k \sum_{i=1}^n |x_i^{(j)} - m_j|^2 \quad (\text{A.1})$$

Donde  $|x_i^{(j)} - m_j|^2$  es la distancia medida entre un dato  $x_i$  y la media de la agrupación  $m_j$ .

#### Paso 3: Revisar criterios de parada

Se revisan los criterios de parada del algoritmo, este para por iteraciones, o por óptimo global.

#### Paso 4: Actualizar

Cuando los criterios de parada no se cumplen, se debe actualizar el algoritmo para una nueva iteración, esto se hace calculando nuevas medias  $m_i$  para cada agrupación

$$\mathbf{m}_i^{t+1} = \frac{1}{|S_i^{(t)}|} \sum_{\mathbf{x}_j \in S_i^{(t)}} \mathbf{x}_j \quad (\text{A.2})$$

Donde  $\mathbf{x}_j$ , representa el conjunto de datos que queda contenido en el grupo  $S_i^{(t)}$ .

### **Paso 5: Resultados**

Al terminar el proceso iterativo, se genera el libro de códigos, que recopila las observaciones más el respectivo grupo al cual fueron asignadas.

## APÉNDICE B

### SOLUCIÓN A LOS TRES PROBLEMAS BÁSICOS EN LOS HMM

#### Solución al problema de la evaluación

Consiste en calcular la probabilidad de la secuencia de observaciones  $O = O_1, O_2, \dots, O_T$ , dado un modelo  $\lambda$  esto es  $P(O|\lambda)$ . Sin embargo al tenerse dependencia de los estados también, se hace necesario definir una secuencia de estados  $Q$ , la cual relaciona el cambio de estados a través del tiempo.

$$Q = q_1, q_2, \dots, q_T \quad (\text{B.1})$$

Donde  $q_1$  es el estado inicial. La probabilidad de la secuencia de las observaciones  $O$  dada una secuencia de estados es

$$P(O|Q, \lambda) = \prod_{t=1}^T P(O_t|q_t, \lambda) \quad (\text{B.2})$$

Se asume que las observaciones son independientes, por lo tanto se tiene la siguiente expresión

$$P(O|Q, \lambda) = b_{q_1}(O_1), b_{q_2}(O_2), \dots, b_{q_T}(O_T) \quad (\text{B.3})$$

Donde  $b_{q_1}(O_1), b_{q_2}(O_2), \dots, b_{q_T}(O_T)$ , son las probabilidades correspondientes a las observaciones dado un estado y un modelo. Para darle solución a la ecuación 3.11, se recurre al procedimiento iterativo de *Forward-Backward*, primero se define la variable de forward como se muestra en la ecuación B.4.

$$\alpha_t(i) = P(q_1, q_2, \dots, q_t = S_i | \lambda) \quad (\text{B.4})$$

El algoritmo para resolver esta ecuación es el siguiente:

Paso 1. Inicialización

$$\alpha_t(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

Paso 2. Inducción

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{i,j} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N \quad (\text{B.5})$$

Paso 3. Terminación

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (\text{B.6})$$

De una manera similar se define la variable *Backward* la cual se muestra en la ecuación B.7. Al igual que la anterior variable servirá para escoger el mejor camino en este caso es el camino de regreso.

$$\beta_t(i) = P(q_1, q_2, \dots, q_t = S_i | \lambda) \quad (\text{B.7})$$

Esta ecuación se puede resolver utilizando el siguiente algoritmo:

Paso 1. Inicialización

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (\text{B.8})$$

Paso 2. Inducción

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1 \quad 1 \leq i \leq N \quad (\text{B.9})$$

Para el proceso iterativo por *Backward*, se requiere  $N^2T$  cálculos y su finalización es de manera similar a la del algoritmo de *forward*.

### Solución al problema de la decodificación

A diferencia del problema anterior, el problema de la decodificación se puede resolver por múltiples caminos, esto con el fin de encontrar la secuencia de estados óptima asociada con la secuencia de observaciones dada. La dificultad está en encontrar los criterios de optimización precisos. Sin embargo uno de los criterios de más peso es encontrar la secuencia de estados que maximice  $P(Q|O, \lambda)$ , lo cual resulta equivalente a maximizar  $P(Q, O|\lambda)$ , una técnica formal existente para solucionar este problema, está basada en métodos de programación dinámica, este método es llamado el algoritmo de Viterbi.

El algoritmo de Viterbi [30, 31] tiene como tarea encontrar la mejor secuencia de estados  $Q = \{q_1, q_2, \dots, q_T\}$ , para optimizar la probabilidad de la secuencia de observación dada  $O = \{O_1, O_2, \dots, O_T\}$ .

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_t} P[q_1, q_2, \dots, q_t = i, O_1 O_2 \dots O_t | \lambda] \quad (\text{B.10})$$

Para la expresión de la ecuación B.10, al aplicarse la programación dinámica, se cuenta con un único camino, el cual tiene la mejor puntuación (esto se refiere a la mejor probabilidad), por inducción se tiene entonces la ecuación B.11.

$$\delta_{t+1}(j) = \left[ \max_i \delta_t(i) a_{ij} \right] b_j(O_{t+1}) \quad (\text{B.11})$$



Para mantener un seguimiento ordenado de la secuencia de estados a través de la programación dinámica, se hace necesario recurrir a una matriz de seguimiento  $\psi_t(j)$ . En la cual se guardaran para cada estado en  $t$  su respectivo valor en  $j$ .

El algoritmo para dar solución a B.11 se muestra a continuación.

Paso 1. Inicialización

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (\text{B.12a})$$

$$\psi_1(i) = 0. \quad (\text{B.12b})$$

Paso 2. Recursión

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T \quad 1 \leq j \leq N \quad (\text{B.13a})$$

$$\psi_t(j) = \arg\max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T \quad 1 \leq j \leq N \quad (\text{B.13b})$$

Paso 3. Terminación

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (\text{B.14a})$$

$$q_T^* = \arg\max_{1 \leq i \leq N} [\delta_T(i)] \quad (\text{B.14b})$$

Paso 4. Camino de vuelta hacia atrás (secuencia de estados)

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T - 1, T - 2, \dots, 1 \quad (\text{B.15})$$

### Solución a la tarea del entrenamiento

La solución al tercer problema es la más difícil, esta consiste en determinar un método para ajustar los parámetros del modelo  $\lambda = (A, B, \pi)$ , de tal manera que se pueda maximizar la probabilidad de que la secuencia de observaciones dada venga del modelo  $P(O|\lambda)$ . Para este problema no se tiene una solución analítica, por lo cual se recurre a encontrar un óptimo local que maximice  $P(O|\lambda)$ . El procedimiento iterativo utilizado es conocido como el método de Baum-Welch o el método EM (Expectation Maximization) [6,32].

A continuación se describe el método iterativo de Baum-Welch o de máxima expectativa.

Con el fin de describir el procedimiento de re-estimación para los parámetros de los HMM, primero se define  $\xi_t(i, j)$ , que corresponde a la probabilidad de empezar en el estado  $S_i$  en el tiempo  $t$ , y llegar al estado  $S_j$  en el tiempo  $t + 1$ .

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (\text{B.16})$$

Igual que para los casos anteriores se puede por medio de inducción estimar las variables de forward y Backward, las cuales quedan escritas de la siguiente forma (ecuación B.17).

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | \lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_i^N \sum_j^N \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (\text{B.17})$$

Ahora se define la variable  $\gamma_t(i)$  como la probabilidad de empezar en el estado  $S_i$ , en el tiempo  $t$ , dada una secuencia de observaciones y un modelo. Por lo tanto se puede relacionar  $\gamma_t(i)$  con  $\xi_t(i, j)$  por una sumatoria sobre  $j$ , esta viene dada

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (\text{B.18})$$

Si sumamos  $\gamma_t(i)$  para todo el conjunto de tiempo indexado a  $t$ , tendremos el conteo de la cantidad de veces que pasamos por  $S_i$ . Otra manera de interpretar estos valores se muestra a continuación [6].

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{número esperado de transiciones para } S_i \quad (\text{B.19})$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{número esperado de transiciones del estado } S_i \text{ al estado } S_j \quad (\text{B.20})$$

Usando las anteriores fórmulas, se puede entonces plantear el método de re-estimación para los parámetros del HMM. Estas se muestran en la ecuación B.21-B.23.

$$\bar{\pi}_i = \text{Frecuencia esperada en el estado } S_i \text{ en } (t = 1) = \gamma_1(i) \quad (\text{B.21})$$

$$\bar{a}_{ij} = \frac{\text{número de transiciones esperado del estado } S_i \text{ al } S_j}{\text{número de transiciones esperadas en el estado } S_i} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (\text{B.22})$$

$$\bar{b}_j(k) = \frac{\text{número esperado de veces que el simbolo } V_k \text{ esta en el estado } j}{\text{número esperado de veces en el estado } j} = \frac{\sum_{t=1}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(j)} \quad (\text{B.23})$$

Con las ecuaciones (B.21)-(B.23) se calcula el nuevo modelo re-estimado  $\bar{\lambda}$  al cual evaluaremos su desempeño mediante la función auxiliar (ecuación B.24).

$$Q(\lambda, \bar{\lambda}) = \sum_Q P(Q|O, \lambda) \log[P(O, Q|\bar{\lambda})] \quad (\text{B.24})$$

Una manera de resolver la función auxiliar de desempeño es mediante los multiplicadores de LaGrange, los cuales aplicamos a  $P(O|\lambda)$ , de donde obtenemos la máxima probabilidad de la secuencia dada un modelo  $\max P(O|\lambda)$ , cuando:

$$\pi_i = \frac{\pi_i \frac{\partial P}{\partial \pi_i}}{\sum_{k=1}^N \pi_k \frac{\partial P}{\partial \pi_k}} \quad (\text{B.25})$$

$$a_{ij} = \frac{a_{ij} \frac{\partial P}{\partial a_{ij}}}{\sum_{k=1}^N a_{ik} \frac{\partial P}{\partial a_{ik}}} \quad (\text{B.26})$$

$$b_j(k) = \frac{b_j(k) \frac{\partial P}{\partial b_j(k)}}{\sum_{l=1}^M b_j(l) \frac{\partial P}{\partial b_j(l)}} \quad (\text{B.27})$$

En cada nueva iteración que se tenga, siempre se deben revisar los criterios de valides y convergencia para el modelo, como sigue

$$\sum_{i=1}^N \bar{\pi}_i = 1 \quad (\text{B.28})$$

$$\sum_{j=1}^N \bar{a}_{ij} = 1, \quad 1 \leq i \leq N \quad (\text{B.29})$$

$$\sum_{k=1}^M \bar{b}_j(k) = 1, \quad 1 \leq j \leq M \quad (\text{B.30})$$